Università
Ca'Foscari
Venezia

Doctoral Program
in Linguistics
Cycle 34

Doctoral Thesis

Memory retrieval and implicit prosody in reading: anaphora resolution by L1 and L2 speakers of English

Coordinator of the Doctoral Program
Prof. Enric Bou Maqueda

Supervisor
Prof. Giulia Bencini

Co-supervisor
Prof. Van-Trao Nguyen

Ph.D. candidate
Thuy-Duong Nguyen
Student ID: 956464

Acknowledgments

My doctoral journey is a journey of back-and-forth traveling between the two continents. Some parts of it were during the Covid pandemic. Still, this once-in-a-lifetime experience is not in any way tedious, thanks to the love and tremendous support from my loved ones and from the extraordinary individuals I have been lucky to be around.

First of all, I am deeply indebted to my dissertation supervisor, Prof. Giulia Bencini, for her invaluable guidance and encouragement during the past four years. I have found myself attracted to psycholinguistics and will continue being so for the years to come, all thanks to Prof. Bencini. She is a knowledgeable mentor who truly inspires all the mentees working with her, myself included. Her passion and dedication in teaching and mentorship are good indicators of how a respected and competent teacher-researcher should be. She always returned my works with detailed comments and constructive feedback. Whenever I sat down with her, listening to her encouraging words, I clearly understood why I had started this journey. Because of her passion for research and consideration for students, I have never second-guessed the path I chose to be on. The past year was the most important and, at the same time, most challenging phase of my research, but Prof. Bencini somehow made the last year a lot less intimidating to me. Her inspiration helped me get through the past four years in the most miraculous way.

My gratitude also goes to my co-supervisor - Prof. Nguyen Van Trao. Prof. Trao is my mentor and my senior colleague at Hanoi University. As the university president, Prof. Trao is often swamped with work, but his feedback on my works always came promptly over the weekend. I admire his dedication to work and his knowledge of linguistics and language pedagogy. Prof. Trao has proved to me that one can excel not only in academic work but also in anything one puts effort into. My four years of engaging in the Ph.D. program and working full-

time in the English Department - Hanoi University could not have been this smooth sailing without the help and support of Prof. Trao. And for that, I am forever grateful.

I would also like to thank my committee members: Prof. Eva Fernández, Prof. Valentina Bambini, and Prof. Roland Hinterholz for their constructive comments and feedback on my thesis. Having a committee of renowned and distinguished researchers in the field is truly a privilege. I am particularly grateful to Prof. Hinterholz for his advice during my first semester in the doctoral program.

Getting to know the members of the weekly lab meetings organized by Prof. Bencini is also a blessing to me. My special thanks extend to Prof. Virginia Valian whose comments and feedback helped shape the crucial component of my research. Words fail to express my thankfulness to my doctoral fellow - Michaela Vann - for her patiently sitting down with me for hours to go over my endless list of experimental stimuli, for our now-and-then Whatsapp chats and Google Meets over the past years. Without Michaela's help, I would not have gathered enough native speakers for the first norming survey. I also wish to thank other members of the lab meetings: Ilaria Venagli, Martina Pucci, and others who listened attentively to my presentations at the lab meetings and gave me meaningful questions and helpful feedback.

My appreciation also goes out to Prof. Enric Bou Maqueda and Prof. Marina Buzzoni for their professionalism in assisting international doctoral students like myself. Prof. Maqueda's instant responses to my every email made me feel that I was cared for, and that my voice was heard. I realized that whenever I needed administrative assistance or support, I could always turn to Prof. Maqueda.

The data collection of my research would have been a laborious process if it had not been for the help of my colleagues and students in the English Department - Hanoi University.

My sincere thanks especially extend to my three senior colleagues in the Managing Board of the English Department: Dr. Doan Quang Trung, Dr. Nguyen Quang Vinh, and Dr. Ngo Van Giang for the spiritual support, and for the share of workload during the months I was abroad for my Ph.D. study.

My research is a quantitative study, but working with numbers has never been my cup of tea. I managed to get to the point where I am now - comfortably working with R codes and autonomously writing codes for the linear mixed-effects models of my study - all thanks to my R tutor - Giang Pham. From our first talk at All-day Coffee, I immediately realized that Giang is a young and brilliant statistic analyst. The completion of data analysis would not have been possible without Giang's help. He also knew how to pump up my confidence throughout the entire data analysis process, despite moments of me feeling down. Working with him was enjoyable and very much "statistically" meaningful.

The add-on of doing research, to me, is to have ampler opportunity for appreciation of personal relationships. I would like to express my thankfulness to my long-time friend - Hoai Tran Rogers - for her enthusiasm in helping me to recruit native speakers for the second norming survey of my research. I was lucky enough to get to know Hoai when I was in Ohio, USA for my M.A study more than a decade ago. Though we have a long-distance friendship, Hoai has shown me that good friends are always there whenever we are in need of help.

My Ph.D. life overseas was bearable thanks to Thuy Nguyen - my best friend in college and now my colleague, perfume buddy, and Ph.D. peer. Our late-night talks, hard laughs, and after-meal strolls made the Ph.D. roller coaster ride a lot easier.

Finally, my deepest gratitude goes to my parents, my parents-in-law, my husband, and my two kids. Living away from home, particularly during the third wave of the Covid

pandemic, was mentally difficult for me. I could not have had enough strength and courage to get through the hardest times if it had not been for the encouragement from my parents. Worries have never been vocalized, only supporting smiles and silent sacrifice - my parents are the most incredible life teachers I could ever ask for. While I was away for my Ph.D. study, my husband and our parents took care of my kids, picking them up from school, preparing them fresh home-cooked meals, teaching them new things, taking them out to their favorite places, and most importantly, keeping them safe and sound. I am immensely grateful to my husband - a man of a few words who knows how to express love and feelings for loved ones through actions; and my two kids - the love of my life - for constantly warming my heart with cute drawings, funny stories, and touching words. I do what I do - all because of them.

MEMORY RETRIEVAL AND IMPLICIT PROSODY IN READING: ANAPHORA RESOLUTION BY L1 AND L2 SPEAKERS OF ENGLISH

Abstract

This thesis focuses on L1 (native English speakers) and L2 (Vietnamese speakers of English) processing and resolution of reflexive-antecedent dependencies. The experimental design crossed the three factors: (i) implicit prosody in the form of text presentation formats (*word-by-word, phrase-segment,* and *whole-sentence*); (ii) syntactic structure (*simple, complex*); and (iii) grammaticality (*target-match, target-mismatch*). An example of the stimuli is shown in (I).

(I) The actress that {Mary/John} interviewed at the awards ceremony {about two years ago/held outside the theater} described {herself/himself} as an extreme workaholic.

Participants were asked to (i) read the experimental sentences presented on the computer screen in one of the three reading formats: word-by-word (*rapid serial visual presentation – RSVP*), phrase-segment (*self-paced reading*), and whole-sentence (*self-paced reading*), then (ii) complete the two follow-up tasks: grammaticality judgment and comprehension probes. Participants' English reading proficiency was measured via a reading test of which the scores significantly and positively correlated with the results of the self-rated proficiency questionnaire.

During sentence processing, participants are expected to employ a retrieval probe that is created by the combination of retrieval cues in order to integrate previously accessed input with current linguistic representations (Martin, 2018). Retrieval of cues during processing might be obstructed by interfering items that share linguistic features with those of the target (e.g.,

+FEM[1]: *actress, Mary, herself* as in (I)). Cue overload occurs when multiple items have features that match retrieval cues, resulting in longer retrieval latency and higher probability of misretrieval (Engelmann, Lewis, & Vasishth, 2019; Lewis & Vasishth, 2005; Parker, 2019; Parker, Shvartsman, & Van Dyke, 2017). When a comprehender fails to deploy retrieval cues due to heavy cognitive load (e.g., cue confusion, increased syntactic complexity, disruption in prosodic contours, etc.), the comprehender is predicted to resort to a processing strategy that is "good enough" to maintain general comprehension (Ferreira, Bailey, & Ferraro, 2002; Ferreira, 2003; Ferreira & Patson, 2007). Sacrifice of detailed syntactic analysis in exchange for heuristic computation of input is suggested to result from the language system's preference for a cognitive equilibrium state that prioritizes early arrival at comprehension (Karimi & Ferreira, 2016). Though prosodic parsing that is in line with the syntactic structure of a sentence is expected to reduce cognitive load, thus increasing probability of accurate retrieval (Bader, 1998; Fodor, 1998, 2002), no prosodic information seems to provide any further help with syntactic processing once the comprehender opts for a simple, good enough processing approach. Language proficiency in general and reading fluency in particular are predicted to mediate between utility of prosodic information and deployment of cues during retrieval.

Results of the study showed that there was evidence of trade-off effects between general comprehension and grammaticality judgment. Interference effects across experimental items were experienced by both groups of participants, particularly among L2 participants. The effects were found to be most robust when there was an interaction between grammaticality and syntactic complexity (i.e., in complex and target-mismatch configuration). Though both L1 and L2 participants benefited from the *sentence* format across experimental tasks, the two groups

[1] FEM: feminine (gender feature)

differed in the effects of the other two reading paradigms. L1 speakers were most disrupted by the *word-by-word* format, and L2 speakers by the *phrase-segment* presentation format. I suggest that the differences between the two groups be attributed to the differences between L1 and L2 processing strategies and the development of reading fluency. Overall, the findings of the study contribute to the processing model based on the framework of the Good-enough (Ferreira et al., 2002; Ferreira, 2003; Ferreira & Patson, 2007) and Online Cognitive Equilibrium (Karimi & Ferreira, 2016) hypotheses. This model takes into account the combined effects of cognitive load, memory access, and implicit prosody (Pratt & Fernández, 2016).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1. Introduction

During sentence processing, comprehenders use different sources of linguistic and nonlinguistic information ranging from knowledge of morphemes, words, sounds, to grammar. How those sources of information are lined up systematically during processing depends largely on human cognitive skills and brain functions such as attention span, sensory recognition, memory retrieval, etc. This dissertation looks into L1 and L2 processing of reflexive-antecedent dependencies as shown in (I). Recent research has pointed out that reflexives, though selective, are affected by interference effects due to cue overload during processing (Parker & Phillips, 2017; Patil, Vasishth, & Lewis, 2016). The deployment of both structural (i.e., binding principles) and non-structural cues (i.e., number, gender, animacy) at retrieval leads to increased cross-association between cues and item features (Engelmann, Lewis, & Vasishth, 2019). When cues match multiple features of different items, cognitive load might increase as a result of fan effects and cue confusion.

Under heavy cognitive load, comprehenders are predicted to prioritize a simple, heuristic processing route that is "good enough" to sustain general comprehension (Ferreira, Bailey, & Ferraro, 2002; Ferreira, 2003; Ferreira & Patson, 2007). The impetus behind the Good-enough hypothesis is the Online Cognitive Equilibrium hypothesis (Karimi & Ferreira, 2016). Under Online Cognitive Equilibrium hypothesis, acceleration in cognitive load such as increased syntactic complexity or interference effects will give rise to a greater reliance on a heuristic approach in which interim input is created during processing. The language processing system was suggested to be in favor of a processing route that arrives at early comprehension so that the state of equilibrium can be attained at the earliest opportunity. Once entering the equilibrium state, the processing system chooses to linger in the state for as long as possible

(Karimi & Ferreira, 2016). Only when required otherwise by the task or when assumed necessary by the parser, a syntax-based analytic route will take over. For reflexive-antecedent dependencies, an accurate syntactic analysis is reflected in the accurate retrieval of the target antecedent. For that reason, maintenance of general comprehension in the absence of accurate retrieval might be a good indicator of the comprehender's deviation from a syntactic algorithmic route for the prioritization of a good-enough processing strategy.

In addition, during silent reading, readers also make use of prosodic cues and prosodic information to guide syntactic parsing and resolve syntactic dependencies (Bader, 1998). A large body of earlier studies revolved around the effects of implicit prosody on syntactic disambiguation of relative clause attachments and garden path sentences (e.g., Augurzky, 2006; Bader, 1998; Bradley, Fernández, & Taylor, 2003; Fernández, Bradley, Igoa, & Teira, 2003; Fordo, 1998, 2002; Harris, Jun, & Royer, 2016; Pratt & Fernández, 2016). To date, not much research has been conducted on the integration of implicit prosody in the processing and resolution of anaphors, particularly reflexive anaphors. This dissertation aims to take on that road, specifically looking into how implicit prosody was mapped onto a cognitive framework that accommodates memory access architecture, Good-enough and Cognitive Equilibrium hypotheses in the processing and resolution of reflexive-antecedent dependencies.

I predicted that comprehenders of different language profiles (i.e., L1 and L2 speakers of English) would adopt the prosody-memory integrated model during processing (Pratt & Fernández, 2016), though varying effects might occur among comprehenders due to individual differences such as reading fluency and language proficiency. The motivation for the recruitment of different groups of language speakers in the present study, firstly, stems from the insufficient amount of research on L2 processing of reflexive-antecedent dependencies.

Secondly, the inclusion of both native and nonnative speakers in the study can provide further insight into the predictions of the prosody-memory integrated model (Pratt & Fernández, 2016) which claims to be "applied to all speakers of a language, regardless of language background" (p. 1).

The dissertation is organized into six chapters. In Chapter 2, I review the relevant theoretical concepts and empirical studies on anaphora resolution, the interplay between implicit prosody and syntax, and the memory access architectures in L1 and L2 sentence processing. I explain in detail the methods used for data collection in Chapter 3, including descriptions of experimental design, experimental stimuli, recruitment of participants, and set-up of experiments. The results of the study are presented in Chapter 4. In this Chapter, I present justifications for the selection of mixed models used for data analysis and provide thorough descriptions of results. In Chapter 5, I discuss the findings with respect to the three areas of inquiry: Cue-based Memory Retrieval, Implicit Prosody, Good-enough Processing and Online Cognitive Equilibrium hypotheses. Chapter 6 concludes my research with a summary of main findings and recommendations for avenues of further research.

Chapter 2. Literature Review

This chapter presents the theoretical concepts and relevant studies concerning anaphora resolution, specifically the resolution of reflexive anaphors. In this chapter, the review of L1 and L2 processing of reflexive-antecedent dependencies is organized following the three theoretical accounts: (i) Cue-based Memory Retrieval, (ii) Implicit Prosody, and (iii) Good-enough Processing and Online Cognitive Equilibrium hypotheses.

2.1. Anaphora Resolution

The term *anaphora* has its origin from the Greek word *anapherein* (*ana-* means *back* and *pherein* refers to *to bear*) and is grammatically defined as "the use of a word which refers to, or is a substitute for, a preceding word or group of words" (Simpson & Weiner, 1989, p. 436, as cited in Schmolz, 2015, p. 20).

(1) The ***female participant*** was seated at a height-adjustable table. ***She*** had her head stabilized by a chin-rest.

In (1), *she* is an anaphor referring to the antecedent *the female participant*. Anaphora, thereby, describes the relationship between an anaphor and its antecedent. The process of identifying the antecedent that correctly matches an anaphor is referred to as *anaphora/anaphor resolution* (Schmolz, 2015), or *anaphora comprehension* (Autry & Levine, 2014).

With respect to the form of anaphors and antecedents, an anaphor can be in the form of a word, a phrase, or a gap as in the case of ellipses:

(2) He asked me to go with him, but I do not want to____. (Schmolz, 2015, p. 150)

An antecedent can be a word, a phrase, a clause, or a combination of clauses or sentences. Anaphors are syntactically classified into *intrasentential* anaphors (when an anaphor

and its corresponding antecedent(s) are positioned within the same sentence) and *intersentential* anaphors (when an anaphor and its antecedent(s) are in different sentences).

2.1.1. Fundamentals of Anaphoric Relations

There are different types of anaphoric relations, among which are *coreference* and *substitution*. The former represents the "prototypical and simplest anaphoric relation" (Schmolz, 2015, p. 23). As in (1), the anaphor *she* is semantically associated with the antecedent *the female participant*, and thus the two items are coreferential.

(3) The green ***apple*** looked tempting, but she chose the red ***one***.

Example (3) demonstrates *substitution* relation in which the anaphor *one* is used to replace *apple*. However, *one* and *apple* are not coreferentially related as the red apple is not identical to the green apple. Substitution does not exclude replacement that requires morphological change of the word(s) (Schmolz, 2015):

(4) The green ***apple*** looked tempting, but she chose all the red ***ones***.

Ones could still be classified as substitution for the antecedent *apple* regardless of the morphological change from singularity *apple* to plurality *ones*.

Halliday and Hasan (2008, pp. 88 – 90, as cited in Schmolz, 2015, p. 24) explained that the difference between coreference and substitution "mainly lies in the linguistic level involved". The anaphor and antecedent in coreference relation are associated on the semantic level, while those in substitution are related on the lexicogrammatical level as respectively shown in (1) and (4).

The difference between coreference and substitution also extends to the replaceability of the lexical items or the expressions. Substitution, by definition, occurs when an anaphor can

replace its antecedent without being necessarily coreferential. However, replacement may not be grammatically feasible in certain cases of coreference:

(5) ***The woman*** told ***herself*** not to do that again.

The replacement of *herself* by *the woman* in the object position ("***The woman*** *told* ***the woman*** *not to do that again*.") results in the semantic change of the sentence.

However, the boundary between coreference and substitution is not always clear-cut since there are items or expressions that cross-section between the two such as the following cases of independent possessive pronouns:

(6) ***Jane*** and ***John*** both brought some cakes to the party today. While ***her*** cakes are mostly cupcakes, ***his*** is a gigantic red velvet cake with mouth-watering cream cheese frosting.

The anaphor *his* refers to both *John* and *John's cake*, thus is classified as an item belonging to the miscellaneous category that shows both coreferential and substitutional relations.

Linguists and researchers have proposed different classifications of anaphors based on either the form of the anaphor (e.g., Schmolz, 2015), the form of the antecedent, the positions of the anaphor and the antecedent, the syntactic functions of the anaphor and the antecedent (e.g., Huang, 2000), the relation between the anaphor and the antecedent (e.g., Quirk, Greenbaum, Leech, & Svartvik, 2012), or the computational aspects of anaphora resolution (e.g., Mitkov, 2002), etc. Following the categorization of anaphors suggested by Schmolz (2015) which focuses on the form of anaphors, there are 12 different types of anaphors (p. 46):

- Central pronouns;
- Reciprocal pronouns;

- Demonstrative pronouns;
- Relative pronouns;
- Adverbs;
- Noun phrases with *the*;
- Proper names;
- Indefinite pronouns;
- Other forms of coreference and substitution;
- Verb phrases with *do*;
- Ellipses;
- Non-finite clauses

Quirk et al. (2012) and Schmolz (2015) grouped three types of pronouns: personal, possessive, and reflexive pronouns into one single category – *central pronouns* – as the three types of pronouns share common grammatical features that are more salient than those shared by the other types of pronouns. Regarding the frequency of usage, central pronouns are documented to occur most frequently in corpus (Quirk et al., 2012).

A reflexive pronoun, per definition, "reflects another nominal element of the clause or sentence, usually the subject, with which it is in a coreferential relation" (Schmolz, 2015, p. 66). Apart from the referential function as in (5), reflexive pronouns are also used with emphatic purposes as in (7).

(7) (a) Sarah ***herself*** could carry this heavy bag upstairs.

(b) Sarah could carry this heavy bag upstairs ***herself***.

The reflexive pronoun *herself* used in the two sentences in examples (7a) and (7b) shows the possibility of reflexive pronouns to occupy different syntactic positions in a sentence for emphatic purposes.

The two possible uses of reflexive pronouns and personal pronouns in object position in constructions such as (8a) and (8b) show the distinction between the two types of pronouns. While *himself* refers to the subject *Jack*, *him* implies another person that is not mentioned in the provided sentence.

(8) (a) Jack told ***himself***.

(b) Jack told ***him***.

Apart from anaphoric reference, the three types of central pronouns all have cataphoric use though restricted in certain instances as shown in the following examples:

(9) Though ***they*** wanted to stay up for another story, ***the children*** decided to go to bed early.

(10) During ***their*** vacation, ***the couples*** went to lots of historical places.

(11) (For) ***himself***, ***Peter*** bought this fancy villa.

The central pronouns functioning as anaphors mostly have coreferential relation with their antecedents, though substitutional relation is also found in constructions involving the use of independent possessive pronouns as in (6).

2.1.2. Theoretical Accounts of Anaphora Resolution

As the theoretical grounds of anaphora processing and resolution extend to multidisciplinary areas (Branco, McEnery, & Mitkov, 2005), an exhaustive review of studies across multiple disciplines is beyond the scope of this research. Instead, this section makes an

attempt to scrutinize the central theories revolving around anaphora resolution with an emphasis on the resolution of anaphoric reflexives.

2.1.2.1. Syntactic Constraints in Anaphora Resolution

The processing of an anaphor initiates upon the first reading of the anaphoric item and is normally completed after the comprehender gathers adequate syntactic and morphological information provided in the context. Subsequent reading of the materials in combination with retrieval of past information are expected to contribute to the interpretation of the anaphor. This section discusses the syntactic relation between an anaphor and its antecedent(s) on the ground of Chomsky's (1981, 1986) Binding Theory. Memory retrieval – another factor contributing to the processing of reflexive-antecedent dependencies – will be reviewed in detail in section 2.1.2.3.

Syntactically, the resolution of anaphors depends on the binding properties of the anaphor in relation to its antecedent (Pollard & Sag, 1992). Binding theory (Binding Principle A, Binding Principle B, and Binding Principle C) is one of the two sub-theories under Chomsky's (1981) Government and Binding Theory which deals with the interpretation of nominal phrases and their indexing relationships (Gardelle, 2012). The concepts of co-indexation and binding refer to the coreferential relationship between a nominal phrase and its anaphoric expression which is in the form of another nominal phrase in the sentence. A noun phrase is bound with another noun phrase under the condition that the two phrases are associated in terms of meaning. Under the generative framework, syntactic anaphora differs from discourse anaphora. The framework does not consider discourse anaphora as *anaphora* since it is not governed by any grammatical relations as the syntactic anaphora is.

Section 2.1.1 above lists out 12 types of anaphors on the level of both inter- and intra-sentential anaphors. With respect to Binding Theory, "the study of anaphora is restricted to nominal expressions in same-sentence uses" (Harbert, 1995, p. 179, as cited in Gardelle, 2012, p. 26). Henceforth, intra-sentential anaphors are irrelevant under the theoretical tenets of Binding Theory, and therefore, are not discussed here.

(12) ***Anna*** poured ***herself*** a cup of tea then sat down on the floor.

Binding Principle A syntactically allows *herself* to be the legitimate anaphor of the antecedent *Anna* for the following reasons: *Anna* and *herself* have a coreferential relationship; and *Anna* precedes *herself* so that *herself* is bound by *Anna* in their local syntactic domain (i.e., their governing category).

C-command (constituent-command) and co-indexation are the two required conditions for binding. Haegeman (1991, p. 198) provided a definition of binding constraint as follows:

- A binds B if and only if A c-commands B;
- and A is co-indexed with B.

One of the earliest concepts of c-command suggested by Reinhart (1976) is used to refer to the relationships between different nodes in a grammatical parse tree. The definition of c-command established by Reinhart (1976, p. 8) indicates that:

- A does not dominate B;
- B does not dominate A;
- and the first (i.e., lowest) branching node that dominates A also dominates B.

The first clause in example (12) can be analyzed into constituents as follows:

[IP [NP_1 *Anna*] [VP [V *poured*] [NP_2 *herself*] [NP_3 [DET *a* [NP_4 *cup*]] [PP [PREP *of*] [NP_5 *tea*]]]

The initial noun phrase (NP_1) *Anna* in example (12) c-commands the second noun phrase (NP_2) *herself* because NP_1 and NP_2 do not dominate one another, but the first/lowest node (the inflectional phrase – IP) that dominates NP_1 also dominates NP_2.

In regard to co-indexation, *Anna* in (12) is co-indexed with *herself* because both of the two NPs refer to the same person/entity.

Reinhart (1983, as cited in Gardelle, 2012, p. 28) further elaborated on a c-command constraint on anaphora:

- A pronoun (whether a pronominal or an anaphor) cannot c-command its antecedent (whereas an antecedent might c-command a pronoun);
- Two lexical NPs cannot co-refer if one c-commands the other.

Taking into consideration the definitions of binding, and c-command constraints, as previously stated, *Anna* is the antecedent of *herself* in (12) because *Anna* binds *herself* and is co-indexed with *herself* in their local syntactic domain. The movement of *herself* to the syntactic position of *her* and vice versa would result in an ungrammatical sentence (**Herself poured Anna a cup of tea then sat down on the floor.*) because a reflexive pronoun c-commands its antecedent within its local/binding domain (Binding Principle A).

Anaphors, under Binding Constraint theory (specifically Binding Principle A), only include reciprocals and complement reflexives as in (12) and (13) (in opposition to emphatic reflexives: ***I** can do this **myself**!*) (Gardelle, 2012). Nonetheless, not all complement reflexives can function as anaphors. In instances where reflexives are used as *override reflexives* – an eligible substitution of a pronoun such as *the whole team and myself* instead of *the whole team and me,* the reflexive *myself* is not regarded as an anaphor (Huddleston & Pullum, 2002, p. 1485; as cited in Gardelle, 2012, p. 29).

As Binding Principle A only allows a reflexive to have its antecedent within its local binding domain (Cunnings & Sturt, 2014), in cases where there are multiple nominal phrases preceding a reflexive as in (13), only the nominal that is within the local domain of the reflexive can be considered the grammatically accessible antecedent.

(13) ***Kim***$_{i}$ suggested *Anna* pour ****herself***$_{i}$ a cup of tea.

(14) *Kim* suggested ***Anna***$_{i}$ pour ***herself***$_{i}$***/her*** a cup of tea.

In example (13), *Kim* is beyond the local domain of the reflexive *herself,* which makes *Anna* – the nominal that is in the local domain – the only possible antecedent of the anaphoric reflexive *herself.*

With respect to the circumstances in which both the reflexive and the personal pronoun can semantically compete for the syntactic position of an anaphor as in (14), both Binding Principles A and B are considered for the interpretation of the anaphor. While Principle A accepts only *herself* to be the legitimate anaphor of the antecedent *Anna,* Principle B allows a pronoun to be beyond the local domain. Thus, Principle B accepts *her* as an anaphor of any nominal phrases that could be syntactically placed outside its local domain. In this case, the antecedent of the anaphoric pronoun *her* could be *Kim* or any female person that is mentioned earlier in the context.

Reflexives and personal pronouns, as suggested by Binding theory, are in complementary distribution. However, there also exist other cases which challenged this complementarity, and even claimed to falsify Chomsky's Binding theory (e.g., Pollard & Sag, 1992; Reinhart & Reuland, 1993; Runner & Kaiser, 2005). Pollard and Sag (1992) confirmed that "non-subject coargument anaphors are the only anaphors that should be constrained by

Principle A" (p. 12), and there are exempt instances of anaphors that are not subject to Principle A. Coargumenthood and exempt anaphors are discussed below.

(15) ***John*** saw that there was a picture of ***himself*** by the fireplace.

Example (15) is one of the instances that runs counter to Binding Principle A as the antecedent *John* of the anaphoric reflexive *himself* is not syntactically positioned within the local domain with its anaphor, but in a higher position in the clause hierarchy. Instances such as (15) consequently gave rise to the reformulation of Binding theory (Chomsky, 1986). The revised Binding theory extended the syntactic domain of a reflexive whose antecedent is not within its most local domain to the domain that might be higher in the clause as in (15). Reluctance to accept this reformulation of Binding theory is reflected in the establishment of the Coargument Relationship theory (Pollard & Sag, 1992; Reinhart, 1983; Reinhart & Reuland, 1993; Reuland, 2001b, 2011; as cited in Cunnings & Sturt, 2014, p. 118). The theory defines coargumenthood as the relationship between arguments of a predicate. In cases such as (12), the verbal predicate *poured herself a cup of tea* requires two coarguments: (i) the anaphoric reflexive *herself* functioning as the object, and (ii) the antecedent *Anna* as the subject of the sentence. Predictions of Binding theory (i.e., complementary distribution of reflexives and pronouns) are plausible in this case. In contrast, in (15), *a picture of himself* is a nominal predicate in which the reflexive *himself* does not have any coarguments within its local domain (the nominal predicate), which clashes with the predictions of Binding theory, specifically the complementary distribution of reflexives and pronouns. An anaphor that takes a remote antecedent beyond the domain that contains the anaphor is called *exempt* anaphor (Pollard & Sag, 1992). Others gravitated towards a more rigorous departure from the term *anaphor* and treated those instances as coargument reflexives (e.g., Reuland, 2001b, 2011).

Concerning referential noun phrases with a possessor like example (16), it has been generally agreed that referential NPs with a possessor are sensitive to binding constraints on the account that the reflexive must be locally bound to its antecedent (which is a possessor) while the pronoun is free from that nominal domain.

(16) ***John's***$_i$ description of ***himself***$_i$/****him***$_i$ was flawless. (Pollard & Sag, 1992, p. 6)

Nevertheless, visual world paradigm studies on referential NPs with a possessor suggested a counter-argument to Binding Constraint theory (e.g. Runner, Sussman, & Tanenhaus, 2003, 2006).

(17) Look at ***Ken***. Have ***Joe*** touched ***Harry*****'s** picture of ***himself***.

Runner et al. (2003, 2006) monitored participants' eye movements while the participants were reading experimental stimuli as (17). Runner et al. (2003, 2006) found that the participants had a tendency to fixate on the possessor NP *Harry* or the subject *Joe* which are syntactically in the same sentence, rather than the nominal *Ken* in the introductory sentence. Findings from visual world paradigm experiments such as Runner et al.'s indicate that reflexives in referential NPs with a possessor should also be categorized as exempt anaphors which are not only restricted by binding constraints, but are also sensitive to other constraints that guide anaphora interpretation such as discourse information.

A more recent approach to Binding theory presented in Sportiche's (2013) research, on the other hand, supported the validity of the conventional Binding Principle A. Charnavel and Sportiche (2016), and Sportiche (2013) argued that conclusions drawn from studies conducted by Pollard and Sag (1992), and Reinhart and Reuland (1993) were inconclusive regarding the exclusion of exempt anaphors from Binding theory, and that though "[a reflexive] can behave as an exempt anaphor, it does not mean that it must. It may well be either" (Sportiche, 2013, p.

202). Examples given by Charnevel and Sportiche (2016), and Sportiche (2013) took into consideration reflexives that have inanimate antecedents. The researchers concluded that a non-personified antecedent locally c-commands its anaphoric reflexive, and that complementary distribution did not exist between anaphors and exempt anaphors. The examples presented in their studies are in French, a language that is linguistically adjacent to English, validating the claim that binding principles guiding the resolution of anaphors are still predicted to have cross-linguistically credibility.

In regard to online processing of anaphoric expressions, structural constraints are not the only constraints that interfere with the resolution and interpretation of anaphors since the nature of online sentence processing requires different processing stages, each of which comes with a wide range of interferential factors. Section 2.1.3 discusses anaphora resolution in online sentence processing in more depth.

2.1.2.2. Non-syntactic Constraints in Anaphora Resolution

Structural constraints are among the factors guiding coreferential processing. In addition to syntactic constraints, the establishment of agreement in non-syntactic features between different nominal phrases helps increase the probability of correct retrieval of antecedent among different candidates.

2.1.2.2.1. Gender

Recent research has shown that both structural (i.e., binding constraints) and nonstructural information (i.e., number, animacy, gender, etc.) are utilized during retrieval of antecedents in anaphora resolution (Parker & Phillips, 2017; Patil et al., 2016). Gender information is critical in early processing of anaphors. A match in gender between retrieval cues and item features was predicted to initiate comprehenders' search for the correct antecedent

(Bock & Miller, 1991; Garrod & Terras, 2000; Parker & Phillips, 2017; Sanford, Garrod, Lucas, & Henderson, 1983; Sturt, 2003). The semantic features of different discourse entities that match the cues are all considered for retrieval as soon as the comprehender encounters an anaphor (Van Berkum, Koornneef, Otten, & Nieuwland, 2007).

Under circumstances where there is only one potential candidate that matches an anaphor in gender, there is a strong possibility that the antecedent will be correctly selected solely based on the matching gender information between the anaphor and the antecedent (McDonald & Macwhinney, 1995). As soon as comprehenders come across *he* in example (18), the gender agreement between *he* and *Tom* immediately suggests that *Tom* is the only correct antecedent in the given context.

(18) ***Tom*** disappointed *Mary* because ***he*** lost his temper at the press conference.

Much debate was circulated among researchers during the 1970s and 1980s (e.g. Caramazza, Grober, Garvey, & Yates, 1977; Smyth & Glencross, 1986; Stevenson & Vitkovitch, 1986; Vonk, 1985a, 1985b) on whether gender information was used during earlier processing. At the time, there was not a consensus concerning the activation of gender cue in first pass reading times. Greene, McKoon, and Ratcliff (1992), and Gernsbacher (1989) used probe paradigms in their studies and both drew similar conclusions that confirmed a restricted use of gender cue in the resolution of pronouns.

Applying similar experimental methods, McDonald and Macwhinney (1995) examined the time course during which gender information is used for anaphora resolution in constructions that contain an implicit causality verb (e.g., *Gary amazed Ellen time after time because he was so talented.*). Three cross-modal probe experiments were conducted with different syntactic constructions manipulated across three experiments. Findings from the

experiments revealed that the processing of anaphors was initiated upon the first encounter of the pronoun. In other words, gender cues facilitated the initial interpretation of pronouns. However, in contrast to the findings from Greene et al.'s (1992) and Gernsbacher's (1989) studies, McDonald and Macwhinney (1995) stated that the onset of anaphora resolution not only depends on the availability of the gender information but also the syntactic structure of the experimental sentences.

Boland, Acker, and Wagner (1998) also observed strong effects of semantic prominence on the resolution of gender-disambiguated pronouns. Ehrlich (1980), Hudson-D'Zmura and Tanenhaus (1998) also reported similar facilitatory effects of gender information on pronoun interpretation, though not during the early stage of processing.

As research investigating the utility of gender information during online processing yielded inconsistent results; plus, there were also concerns that the nature of the probe task might interfere with pronoun resolution, more fine-grained methods were applied to better monitor the effects of gender information. Arnold, Eisenband, Brown-Schmidt, and Trueswell (2000) applied a more sensitive but less invasive experimental method – eye tracking – in their speech processing research. Participants' eye movements were recorded while they were listening to the texts describing the pictures in which pronouns were included to refer to either the first or the second character. Results of the study were consistent with findings from past research, suggesting that gender cues played an important role in guiding the resolution of pronouns, especially during early processing. Gender information was predicted to facilitate processing, particularly when there was an overt demand for the resolution of pronouns (e.g., when there were comprehension questions targeting pronoun-antecedent) (e.g. Garnham, Oakhill, & Cruttenden, 1992).

Kaiser, Runner, Sussman, and Tanenhaus (2009) investigated the interplay between structural and semantic constraints in anaphora resolution using four different experiments which manipulated the inclusion of possessive pronouns in picture-NPs. The overall results from the four experiments indicated that online processing of anaphoric pronouns and reflexives is subject to not only structural constraints but also semantic constraints. Further indications from the findings also pointed to the difference between pronouns and reflexives in sensitivity degrees towards different types of retrieval information. Factors that guide anaphora resolution (e.g., structural and semantic factors) were weighted differently depending on specific types of anaphors. Though the scope of Kaiser et al.'s study was restricted to configurations containing picture NPs with or without a possessor, the findings of the study could be generalized to other constructions that include pronouns and/or reflexives. It was predicted that pronouns would be more susceptible to semantic information and less sensitive to structural constraints than reflexives.

Kreiner, Sturt, and Garrod (2008) examined the effects of stereotypical gender nouns (e.g., *driver*) and definitional gender nouns (e.g., *gentleman*) on the processing and resolution of reflexive anaphors. The researchers used two eye-tracking experiments in which anaphora and cataphora sentences were used respectively in the first and the second experiments. Results from the two experiments indicated that stereotypical nouns and definitional nouns differentially affected the processing and resolution of reflexive-antecedent dependencies.

Prior to Kreiner et al. (2008), a large body of research (e.g., Carreiras, Garnham, Oakhill, & Cain, 1996; Duffy & Keir, 2004; Sturt, 2003) had looked into the effects of stereotypical gender nouns in anaphora resolution. Findings from those studies confirmed the

emergence of reading difficulty when there was a gender clash between an anaphoric reflexive and its stereotypical gender antecedent such as example (19) below.

(19) The ***surgeon*** came into the office and made ***himself/herself*** a cup of coffee.

Reading difficulty was manifested in longer reading time in self-paced reading or longer fixations on the anaphoric expressions in eye-tracking experiments. The term *mismatch cost* was used to refer to reading difficulty (Kreiner et al., 2008) and was claimed to be caused by the disagreement in gender between the stereotypical gender noun and the anaphoric reflexive. As gender is marked at different linguistic levels, gender agreement has certain impact on the interpretation of dependencies "within linguistics levels (e.g., between subject and verb) and between linguistics levels (e.g., the gender of a subject noun and the sex of its referent on a discourse model)" (Kreiner et al., 2008, p. 240). The stereotypical gender noun *surgeon* in (19) exhibits a bias in gender assignment that comprehenders tend to hold towards the noun. Such bias might be so strong that it is always at display, which somehow drives comprehenders to draw certain gender categorical inferences based on their real world knowledge, especially when the discourse information provided is not sufficient enough to guide the interpretation of the referent (e.g., Carreiras et al., 1996; Oakhill, Garnham, & Reynolds, 2005; Reynolds, Garnham, & Oakhill, 2006).

The interpretation of referents including the resolution of anaphors is governed by multiple constraints, some of which require high levels of mental analysis. Garnham (2001) presented the mental models that take into account symbolic and conceptual representations that interfere with the processing of referential expressions. The models suggested that the availability degrees of informative cues given in a context determine the types of inferences that comprehenders could draw from the discourse entities. Concerning example (19), whenever

lexical information fails to guide the interpretation of the referential expression, comprehenders will bring in their real world knowledge (e.g., their perceived categorical gender assignment of *surgeon* as a male person) to compensate for the insufficiency of lexical information. Whether or not such inferences are drawn automatically and subconsciously by comprehenders still provokes controversies among linguists and researchers (e.g., McKoon & Ratcliff, 1992; Carreiras et al., 1996; and Reynolds et al., 2006). The linguistics literature has documented studies that advocated the application of mental models in drawing inferences about gender role nouns during initial processing (e.g., Carreiras et al., 1996). Real world knowledge about stereotypical gender assignment was observed to be activated by comprehenders immediately upon the initial encounter of the role noun. Nonetheless, whenever there is a gender mismatch between morphological information and real world information, comprehenders would draw inferences based on grammatical forms over real world knowledge, which leads to the elimination of the mismatch cost.

Advocates of the automaticity of gender stereotypes argued that gender stereotypes could be treated as lexical features of the gender role nouns. There are certain types of role nouns (e.g., *surgeon, nurse, minister,* etc.) that automatically evoke gender stereotypes (Kreiner et al., 2008). Automatic activation of gender stereotypes was further supported by findings from experimental research that required participants to make either lexical or gender decisions on certain pronouns (e.g., Banaji & Hardin, 1996). Using priming methods, Banaji and Hardin (1996) found that automatic gender stereotyping was experienced in certain role nouns such as *doctor, nurse* via faster responses in instances in which the role noun and the target pronoun agreed in gender (e.g., *doctor* and *he*) compared to when they did not. However, Kreiner et al. (2008) critically presented a number of limitations in Banaji and Hardin's (1996) research.

Firstly, the research did not make a distinction between definitional gender nouns (e.g., *boy, girl*) and stereotypical gender nouns (e.g., *minister, secretary*), which results in a generalized processing mechanism for both types of gender nouns. Secondly, discourse information was neglected in the study as the words used in the tasks were context-free and presented in isolation, indicating that it was the lexical processes that were involved in the gender-priming tasks.

Along with priming experiments, ERP research was also conducted (e.g., Osterhout, Bersick, & Mclaughlin, 1997) to get insight into the differences in the processing of definitional and stereotypical gender nouns. Results from ERP experiments are consistent with findings from Banaji and Hardin's (1996) research, suggesting that both of the two types of nouns had their gender representations activated and processed in a relatively similar manner, and that if there is any processing difference between the two nouns, the difference lies in stereotypical gender nouns. Stereotypical gender nouns are highly probabilistic (i.e., *surgeon* is often associated with a *male* character, while *nurse* is often associated with a *female*), while definitional gender nouns are categorized as either denoting a male or female person (e.g., *man, woman, king, queen*, etc.). The consistency found in the findings of the aforementioned research is partly attributed to the fact that those studies did not take into account discourse-related factors that might also have certain impact on the processing of gender role nouns.

In contrast to the conclusion from Banaji and Hardin's (1996) research, findings from Carreiras et al.'s (1996) study resonated with the predictions from Garnham's (2001) mental models on the account that gender stereotyping does not occur at the lexical level and may also be altered by the given discourse contexts. Thus, the processing of the two types of gender nouns is not identical. With respect to discourse contexts, the processing of stereotypical gender

nouns was also reflected in the Lexical Reinterpretation Model proposed by Hess, Foss, and Carol (1995). The model suggested that when a comprehender first encounters a word, a wide range of information, at least the typical or context independent information, (Barsalou, 1982; as cited in Kreiner et al., 2008, p. 242) related to the word would come into play to help with processing and comprehension. Depending on the given context, such information may be re-evaluated to guide the comprehender in the processing of the word. The Lexical Reinterpretation Model was tested by Duffy and Keir (2004) with a focus on stereotypical gender nouns. Examples such as (20) and (21) were used as experimental stimuli in which the target items are stereotypical gender nouns:

(20) The ***babysitter*** found ***herself/himself*** humming while walking up to the door.

(21) The ***firefighter*** burned ***himself/herself*** while rescuing victims from the building.

Participants were required to read the experimental sentences either in isolation or with provided contexts. Mismatch effect was found in the context-free condition via longer fixation on the reflexive pronoun when the gender of the reflexive conflicted with the target stereotypical role noun. Such an effect indicated immediate activation of gender stereotype upon initial encounter of the target item. In the condition in which the target sentences were disambiguated by a given context, the mismatch effect was eliminated. Findings from Duffy and Keir's (2004) research supported the Lexical Reinterpretation Model suggested by Hess et al. (1995).

Prior to Kreiner et al.'s (2008) research, there was a gap in the literature concerning the differences between stereotypical and definitional gender nouns in language comprehension (e.g., Banaji & Hardin, 1996; Osterhout et al., 1997; Carreiras et al., 1996; Duffy & Keir, 2004). Studies that investigated definitional and stereotypical gender nouns (e.g., Banaji & Hardin,

1996; Osterhout et al., 1997) did not consider the effects of discourse contexts, while those which did (e.g., Carreiras et al., 1996; Duffy & Keir, 2004) failed to take into account the distinctive features of definitional and target gender role nouns (Kreiner et al., 2008). As previously reviewed, Kreiner et al. (2008) examined the effects of discourse information, juxtaposing the two types of gender nouns across experimental sentences. Their research aimed to find out whether discourse information presented prior to the initial presentation of the stereotypical gender noun could affect comprehenders' gender interpretation of the noun, and whether the given preceding context had any influence on the interpretation of definitional nouns to the same extent that it did on that of stereotypical nouns. Overall, findings from their research indicated qualitative differences in the strength of constraints and the processing of the two types of gender nouns. These findings ran counter to findings from a number of previous studies (e.g. Banaji & Hardin, 1996; Osterhout et al., 1997) which claimed that the differences between stereotypical and definitional gender nouns, if any, are quantitative. Kreiner et al. (2008) also confirmed the interaction between linguistic representations (e.g., morphological, syntactic, pragmatic information) and real world knowledge during processing.

Concerning the influence of real world knowledge on language processing, Molinaro, Su, and Carreiras (2016) conducted ERP experiments using Spanish experimental stimuli, and concluded that compared to linguistic factors, social knowledge concerning gender stereotypes affects language processing and comprehension in a different way. The researchers further stated that the influence of syntactic information was modulated by stereotypical knowledge, which consequently raised a question of whether or not gender stereotyping should be treated as a type of semantic knowledge (Molinaro, et al., 2016). However, Molinaro, et al.'s study did not

account for the effects of preceding discourse contexts on reference processing, which may somehow affect the interpretation of the results.

Though there has not been a consensus with respect to the effects, the interplay, and the time course of different linguistic cues in the processing of anaphoric elements, the majority of empirical studies suggested that gender information reduces the possibility of incorrect antecedent assignment, thus facilitating coreference resolution and language comprehension.

2.1.2.2.2. Number and Person

Agreement in number and person between the antecedent and the anaphor was suggested to facilitate the interpretation of the anaphoric expression since semantic information is also conveyed through the number and person feature of the word. However, the notional number of a word is not always congruous with its morphological number. In English, collective nouns (e.g., *group, team*, etc.) are among words which are not usually morphologically marked as plurals, but notionally function as plurals as shown in example (22) (Bock, Eberhard, Cutting, Meyer, & Schriefers, 2001, as cited in Kreiner et al., 2013, p. 830). On the other hand, there are also nouns in English which are marked morphologically as plurals, yet perceived conceptually as singular, for instance: *trousers*, *shorts*, etc.

(22) (a) ***The red team*** considered ***itself*** the winner.

(b) ***The red team*** considered ***themselves*** the winner.

As notional number does not always correspond with morphological number, Corbett (2000, as cited in Kreiner et al., 2013, p. 831), among other linguists, stated that subject-verb agreement is mainly grounded on the grammatical agreement between the subject and the verb, while agreement between an anaphor and its antecedent is predominantly susceptible to conceptual number. Conceptual number dissociates from morphological number as the

processing of anaphor-antecedent and subject-verb dependencies are not homogeneous. Anaphor-antecedent dependencies are subject to *marking* – a conceptually driven process, while subject-verb agreement mainly depends on *morphing* – a grammatically driven process which relies on the syntactic position of a word in a sentence to assign the inflectional affixes that mark number to the word (Bock, Eberhard, & Cutting, 2004, p. 254; as cited in Kreiner et al., 2013, p. 813).

Conceptual and grammatical number in English and in some other languages have been studied in both production and comprehension (e.g., Bock & Eberhard, 1993; Bock et al., 1999, 2004, 2006; Gibson, Pearlmutter, Canseco-Gonzalez, & Hickock, 1996; Haskell & MacDonald, 2003; Kilborn & Cooreman, 1987; Viglioco, Butterworth, & Semenza, 1995; Viglioco, Butterworth, & Garrett, 1996). It was observed in a number of studies (e.g., Bock et al., 1999) that participants had a tendency to add plural markers to an anaphoric pronoun that has a corresponding antecedent in the form of a collective noun (e.g., *the jury for the trial*). When the collective noun is followed by a verb, the verb is more likely to be singularly inflected.

(23) The ***spectator(s)/audience*** at the tennis ***match(es) (behaved/shouted)***. (Bock et al., 1999)

In their experiments, Bock et al. (1999) provided the participants with preambles as shown in example (23) and asked them to continue the sentence with either pronouns or verbs. Findings from the experiments revealed that there was a higher rate of plural markers in case of pronoun completion than verb completion. The explanation given to account for the discrepancy in singular and plural inflection in pronouns and verbs was that the number features of pronouns are conceptually marked at the lexical level during pre-production stage (i.e., planning), while

inflectional affixes that mark singularity or plurality of verbs are determined at a later stage when grammatical information is readily available to the comprehender.

Studies on agreement have mostly been conducted in language production. Since comprehension was the prerequisite for production (Eisner & McQueen, 2005), the two processing mechanisms are interrelated. The process of comprehension requires the comprehender to interpret and, at certain times, decode the given message based on the synthesis of different types of information provided by a variety of linguistic cues. Comprehenders were observed to be relatively alert at detecting agreement violations manipulated in the experimental tasks. Comprehenders' responses to agreement violations were reflected in their performance in different types of experimental tasks: grammaticality judgments (e.g., Blackwell, Bates, & Fisher, 1996; Nicol, Forster, & Veres, 1997), lexical reading speed (e.g., Wagers, Lau, & Philips, 2009), eye-tracking (e.g., Pearlmutter, Garnsey, & Bock, 1999), and ERPs (e.g., Hagoort, Brown, & Groothsen, 1993; Pulvermuller, Shtyrov, Hasting, & Carlyon, 2008).

Nonetheless, comprehenders' sensitivity towards agreement errors varied among studies. Kreiner et al. (2013) investigated the processing of number agreement in sentence comprehension, taking into account experimental instances that included both conceptual and grammatical marking. The rationale for Kreiner et al.'s (2013) study was based on the argument that production-based models (e.g., Bock et al., 1999, 2001, 2004) might not consistently reflect the difference between morphologically inflected words and their conceptual number interpretation in production and comprehension. In the study, Kreiner et al. (2013) compared the effects of conceptual and morphosyntactic number on agreement in both subject-verb agreement and anaphor-antecedent dependencies as shown in example (24).

(24) Decisions about medical issues are often very hard and mistakes can be painful.

The ***assembly/assemblies*** (*panel/panels, housewife/housewives*) convinced themselves that the correct decision was made.

The ***family/families*** definitely and undeniably ***wishes/wish/wished*** to avoid a court trial.

Findings from their eye-tracking experiments were consistent with earlier research on production. Conceptual number differs from morphological number in the effects on the processing of different types of agreement. Anaphor-antecedent agreement is notionally driven, whereas subject-verb agreement is more susceptible to morphosyntactic information. The results provide further evidence to the dissociation between the processing of notional and morphosyntactic features in both sentence comprehension and production. Though there was a lack of consensus over a processing architecture that could fully explain the association between comprehension and production in the processing of number features, Kreiner et al. (2013) suggested that a parallel independent-process architecture, up to the date of their study, was most likely to provide justifications and explanation for the general processing of semantic and notional number in both sentence comprehension and production.

2.1.2.2.3. Discourse Focus

Anaphora resolution is guided by a set of constraints, among which is discourse focus. The interpretation of an anaphor depends on the interplay between a wide range of linguistic factors, for instance, immediate focusing (Garrod & Sanford, 1985; Gordon, Grosz, & Gilliom, 1993), the intersentential assignment of an anaphor to an antecedent as suggested by the parallel assignment strategy (Cowan, 1980), the thematic structure of a given passage in which an anaphor is used to refer to a certain event or character in the text (Cirilo, 1981), and the number

of clauses and sentences in which an anaphoric expression is used (e.g., Ehrlich & Rayner, 1983), etc. The discourse coherence of a given text is determined by the semantic entities in the text. Studies on the relationship between reference and coherence have been guided by Centering theory (Grosz, Joshi and Weinstein, 1983, 1986). Grosz et al. (1986) were among the first to introduce *forward-looking* and *backward-looking* – the two types of discourse centers in an utterance. Centering theory proposed that every utterance (abbreviated as utt_n), except the first utterance in a discourse segment (dis_m), has a backward-looking center (*Cb*) and a set of forward-looking ones (*Cf*) (Grosz et al., 1986). The purpose of the backward-looking center is to connect a particular utterance in a discourse segment with the previous utterances (utt_{n-1}). A set of forward-looking centers, on the other hand, creates the potential connections between the current utterance and the successive ones.

(25) Peter told John about the incident.

Cf: {Peter, John, incident}

(26) He warned him not to tell anyone about it.

Cb: Peter; Cf: {Peter, John, incident}

(27) (a) He asked John if he had any further information about it.

(b) Peter asked him if he had any further information about it.

Cb: Peter; Cf: {Peter, John, incident}

There are two centering rules underlying the Centering theory. Grosz et al. (1986) proposed that the first centering rule dictated the backward-looking center, and that "the most highly ranked element of $Cf(utt_{n-1})$ that is realized in utt_n is the $Cb(utt_n)$ and must be realized by a pronoun if any element of $Cf(utt_{n-1})$ is realized in utt_n by a pronoun" (Grosz et al., 1986; as cited in Gordon et al., 1993, p. 313). In other words, the continuity and coherence of a discourse

segment is realized through the use of a pronoun as the pronoun functions as a link to the nominal mentioned in the preceding discourse.

With respect to the first centering rule, example (25), if assumed to be the initial utterance of a given text, does not have any backward-looking center. Its set of forward-looking centers consists of *Peter*, *John*, and *the incident*. Those three forward-looking centers are orderly ranked following the sequence of appearance in the utterance. The backward-looking center of examples (26) and (27a) is *Peter*, and the most prominent forward-looking center within their set of Cf is also *Peter*. On the contrary, the backward-looking center of example (27b), *Peter*, is not represented by the pronoun *he* as suggested by the first rule, but by the proper name *Peter*. That results in an awkward reading of the utterance, and, to a certain extent, affects the coherence of the discourse.

The second rule deals with relations among the centers across utterances. There are three types of centering relations between two utterances (the current utterance - utt_n and the subsequent one - utt_{n+1}). The degrees of discourse coherence varied among the three types of centering relations. *Continuing* reflects the most coherent transition, followed by *retaining* and *shifting* (Grosz et al., 1986, as cited in Gordon et al., 1993, p. 314):

- Continuing: $Cb(utt_{n+1})$ is the same entity as $Cb(utt_n)$ and is the most highly ranked element of $Cf(utt_{n+1})$.
- Retaining: $Cb(utt_{n+1})$ is the same entity as $Cb(utt_n)$ but some other forward-looking center is more highly ranked in $Cf(utt_{n+1})$.
- Shifting: $Cb(utt_{n+1})$ is different from $Cb(utt_n)$.

Gordon et al. (1993) tested Centering theory by investigating the effects of pronouns on discourse coherence. Results from the self-paced reading experiments showed that the use of a

pronoun as a backward-looking center provides a strong link between the current utterance and the preceding one, which maintains the coherence between discourse segments. Gordon (1993) and his colleagues also confirmed the dissimilarity between backward-looking and forward-looking centers, and further specified that elements within a set of forward-looking centers also differ from one another in terms of prominence.

Sanford, Moar, and Garrod (1988) compared the effects of proper names (e.g., *Jonathan*) and word(s) that describe(s) the role of that person (e.g., *hairdresser*) on the maintenance of coherence of a given text. Findings from their experiments indicated that proper names have a high frequency of usage in subsequent discourse segments, and the use of anaphoric expressions to refer to the previously mentioned character which was represented by a proper name better retains the normal reading of the sentence in comparison with the use of role description words. Sanford et al. (1988) concluded that proper names play an important role in discourse focus.

A named character is highly focused in discourse when it is introduced with a proper name in the first sentence of a given passage, then followed by an anaphoric pronoun in the subsequent discourse segments (Sanford et al., 1988). Discourse focus was also included in studies concerning the role of Binding theory in real time sentence processing (e.g., Sturt, 2003). Review of those studies will be given in section 2.1.3.

Regarding discourse prominence, previous research shows that there are multiple factors that determine the salience degree of a discourse item: referential form, syntactic and semantic function, and coreference repetition (e.g., Garnham, 2001). The more prominent an antecedent is, the more likely it is to be retrieved (e.g., Garnham, 2001; Garrod & Sanford, 1994; Nicol, Fodor, & Swinney, 1994; Nicol & Swinney, 2003). A salient antecedent was predicted to have

more robust activation strength in memory than a less prominent antecedent. There are different approaches to the prominence of discourse items such as the Structure Building framework (Gernsbacher, 1989; Garnham, Traxler, Oakhill, & Gernsbacher, 1996); the Focus Memory framework (Garrod, Freudenthal, & Boyle, 1994; Stewart, Pickering, & Sanford, 2000); and Centering theory (Grosz & Sidner, 1986). The first two approaches viewed activation strength as a continuum which takes into account the dissociation between different types of antecedent focus (i.e., explicit vs. implicit). Following these approaches, higher probability of retrieval was associated with an item's higher degree of prominence. Gundel (1999), and Gundel, Hedberg, and Zacharski (1993) proposed that discourse prominence also contributes to the speed up of the processing of anaphors.

Research on memory retrieval (e.g., McElree, 2001, 2006) also showed that the processing speed of a psychologically focused item increases as compared to that of a less salient item since the prominent item is more actively maintained in focal attention. With the purpose of testing the two different accounts on discourse elements (e.g, Garnham et al., 1996; Gundel, 1999), Foraker and McElree (2007) investigated the effects of prominence on resolution of anaphors. Findings from the experiments showed that the types of experimental constructions governed the maintenance of discourse items in memory. Different types of pronouns were associated with different rates of resolution accuracy: gendered personal pronouns (i.e., *he, she*) were more accurately resolved than the pronoun *it* as the preceding type of pronouns is not as ambiguous as the latter one.

Findings from previous research have pointed to the effects of discourse focus in the resolution of anaphors, especially when an anaphor is represented by a named character introduced earlier in the discourse context. Anaphora resolution was argued to also depend on

the activation strength of the discourse item, which will be discussed in depth in the following section.

2.1.2.3. Cue-based Memory Retrieval in Anaphora Resolution

The anaphor-antecedent relation has been extensively studied over the past decades. Memory access models (e.g., Engelmann et al., 2019; Lewis & Vasishth, 2005; McElree, 2006) have been developed and tested in both language comprehension (e.g., Lewis & Vasishth, 2005; McElree, 2006; Van Dyke & McElree, 2011) and production (e.g., Badecker & Lewis, 2007). As the scope of this current research falls within language comprehension, this section primarily focuses on the review of cue-based memory retrieval in the comprehension and resolution of reflexive-antecedent dependencies, and discusses sentence production whenever relevant.

Under Cue-based Memory Retrieval framework (McElree & Dosher, 1993), both linguistic factors (e.g., syntactic structures, morphological features, etc.) and memory decay are expected to affect processing. Lewis and Vasishth (2005), and Engelmann et al. (2019) suggested that during processing, comprehenders experience immediate memory decay, i.e., as soon as a lexical item is accessed by a comprehender, it will instantly decay in memory (see Christiansen & Chater, 2016). Such decay, nevertheless, is reversed by the reactivation of the item. Activation strength of an item depends on the item's morphosyntactic features. An item hierarchically occupying a more prominent syntactic position or having more salient morphosyntactic features has higher probability of being accessed and reactivated, which results in a higher probability of retrieval (Badecker & Lewis, 2007; Lewis & Vasishth, 2005; Elgenmman et al., 2019).

The rationale for the development of cue-based memory retrieval models is based on the Adaptive Control of Thought – Rational (ACT-R) theory (Anderson & Lebiere, 1998). Under

ACT-R theory, both cognitive (e.g., memory capacity, attention monitoring, etc.) and linguistic constraints (e.g., semantic, syntactic, discourse constraints, etc.) affect language processing. These constraints are reflected in memory decay/rehearsal and (re)activation of a discourse item. Regarding the role of memory, accurate retrieval of an item depends on two major accounts of memory: memory capacity and memory retrieval. Advocates of the former account (e.g., Gibson, 2000; Just & Carpenter, 1992) stated that memory capacity refers to the ability to store and maintain information about a discourse item in memory. Though differing among individuals, memory capacity was argued to affect resolution of discourse dependencies. The second account, in contrast, places emphasis on memory retrieval. In other words, resolution of dependencies was suggested to be dependent on an individual's ability to retrieve the discourse item from memory. Researchers following this perspective on memory (e.g., McElree, Foraker, & Dyer, 2003; Lewis, Vasishth, & Van Dyke, 2006) argued that memory capacity of the human brain is limited and thus does not exclusively determine resolution of dependencies.

Parker, Shvartsman, and Van Dyke (2017) reviewed contemporary research on cue-based memory retrieval in sentence comprehension and argued that "linguistic dependencies are resolved using a direct-access, cue-based retrieval mechanism that gives preferential weighting to syntactic information when navigating linguistic representations in memory" (p. 1). Challenges to the theoretical grounds and the future application of memory retrieval were also discussed in Parker et al.'s (2017) study. Memory retrieval in language processing and comprehension has been substantially studied with a focus on the resolution of dependencies such as ellipsis, subject-verb agreement, and anaphor-antecedent dependencies.

(28) The father ***came into the kitchen***, and the son ***did too***.

(29) ***The researcher*** who rarely sent the doctors to the clinic ***create/creates*** reports for the governor. (Pratt & Fernández, 2016)

(30) ***Jonathan/Jennifer*** remembered that the surgeon had pricked ***himself/herself*** with a used syringe needle. (Sturt, 2003)

Memory retrieval mechanisms were proposed to be driven by either a sequential, serial search (e.g., Sternberg, 1975); a parallel, simultaneous access (e.g., Townsend & Ashby, 1983); or a direct access (McElree, 2000, 206; McElree & Dosher, 1989, 1993) to content in memory. Under the serial search mechanism, each individual item in memory is compared with the target item in a certain sequence. The subsequent comparisons could only begin once the comparison between the previous item and the target is completed (Fific, Nosofsky, & Townsend, 2008). The parallel search mechanism, on the other hand, allows simultaneous comparisons between the target and different items in memory (Townsend & Ashby, 1983). The direct-access mechanism differs from the parallel search in that under parallel models, accurate retrieval depends on the strength of items in memory, while retrieval efficiency, under the direct-access mechanism, depends on the familiarity between items (Parker et al., 2017). Under the direct access mechanism, each item is inspected based on its content rather than its hierarchical position, hence the name "content-addressable" (Parker et al., 2017). The direct access memory retrieval mechanism applies a global rather than a local search of constituents encoded in memory. These encoded items are matched against a retrieval probe which is established based on the contextual and grammatical cues. Probability of misretrieval, as a result, might arise when the retrieval probe matches the features of multiple items. In contrast, the serial search mechanism allows each item to be examined individually in a sequence; thus, interference from distractor items was assumed to be less likely to occur under serial search. However, as each

item in serial search models is inspected sequentially, the search time of the target item might be prolonged, which is not usually observed under the direct access memory mechanism.

Parker et al. (2017) presented a number of concerns regarding Cue-based Memory Retrieval theory in sentence comprehension which according to the authors have not been sufficiently addressed in the literature. Parker and colleagues (2017), stated that cue-based memory retrieval theory still lacks solid theoretical grounds explaining how linguistic constraints associate with retrieval cues during the retrieval process. The theory is also faced with challenges from empirical studies on direct access, and predictive and probabilistic processing. Parker et al. (2017) suggested that *constant time access* (e.g., Martin & McElree, 2008, 2009, 2011; McElree, 2000; McElree, Foraker, & Dyer, 2003) and *interference* (e.g., Fedorenko, Gibson, & Rhode, 2006; Van Dyke, 2007; Van Dyke & Johns, 2012; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006, 2011; Wagers, Lau, & Phillips, 2009) are the strongest sources of empirical evidence for Cue-based Memory Retrieval theory. The evidence of constant time access demonstrated that the retrieval speed of different items in a sentence under the direct access mechanism holds constant across all items regardless of the followings: (i) the syntactic position of the item, (ii) the length of the item, and (iii) the quantity of the distractors presented in the sentence (e.g., Martin & McElree, 2008, 2009, 2011; Oztekin, Davachi, & McElree, 2000; McElree, Foraker, & Dyer, 2003). However, this constant time access is accompanied by a tradeoff between access time and retrieval accuracy, such that misretrieval might occur if there are multiple items that match the retrieval cues.

The other source of evidence points to the two types of similarity-based interference that are likely to occur during processing: *inhibitory interference* and *facilitatory interference*. These two types of interference arise in different contexts and thus, behave differently (Parker et al.,

2017). *Inhibitory interference* occurs when the target item matches other items in certain features, resulting in cue overload at retrieval. Consequently, the target item will be less distinctive or even overwritten by other items in memory (e.g., Van Dyke & McElree, 2006; Van Dyke, 2007). Inhibitory interference effects were reflected in (i) longer reading times at the items that overlap in features, (ii) longer final reading times of the whole sentence, and (iii) higher probability of misretrieval. Examples (31b) and (31d) below illustrate the effects of inhibitory interference (Van Dyke, 2007).

(31) (a) The pilot remembered that the lady who was sitting in the smelly seat yesterday moaned about a refund for the ticket.

(b) The pilot remembered that the lady who was sitting near the smelly man yesterday moaned about a refund for the ticket.

(c) The pilot remembered that the lady who said that the seat was smelly yesterday moaned about a refund for the ticket.

(d) The pilot remembered that the lady who said that the man was smelly yesterday moaned about a refund for the ticket.

Contrastive to *inhibitory interference, facilitatory interference* occurs when the distractor facilitates the processing of a sentence. Such cases have been found in the domain of subject-verb agreement as in (32).

(32) The **key** to the **cell(s)** (unsurprisingly) **was/were** rusty from many years of disuse. (Wagers, Lau, & Phillips, 2009).

The plurality marking on the distractor *cells* matches the plural feature of the verb *were*, resulting in the speed-up of reading time after *were*. Facilitatory interference has also been documented in anaphor-antecedent dependencies (e.g., Parker & Phillips, 2014, 2017; Parker,

Lago, & Phillips, 2015), and negative polarity items (e.g., Drenhaus, Saddy, & Frisch, 2005; Vasishth, Brussow, Lewis, & Drenhaus, 2008; Xiang, Dillon, & Phillips, 2009). Facilitatory interference is responsible for the *illusion of acceptability* phenomenon since the presence of a distractor in the sentence interferes with processing by creating an *illusion* that the ungrammatical sentence is grammatically correct.

Susceptibility to facilitatory interference has also been investigated in the domain of anaphora processing, particularly in the agreement between an anaphoric reflexive and its accessible antecedent. Dillon, Mishler, Sloggett, & Philips (2013) concluded that reflexives are not influenced by facilitatory interference. Similar observations were also reported in a number of other studies (e.g., Clifton, Frazier, & Deevy, 1999; Cunnings & Sturt, 2014; Kush, Lidz, & Phillips, 2015; Nicol & Swinney, 1989; Sturt, 2003). The explanation accounting for the difference between reflexives and subject-verb agreement is that the two types of licensing deploy different sets of cues during retrieval. Nevertheless, the absence of interference effects in the above-mentioned studies might be due to the fact that those studies "may not have used strong enough reflexive-antecedent mismatches to reliably elicit an effect" (Parker & Phillips, 2017, p. 275).

For that reason, Parker and Phillips (2017) manipulated the number and the types of matching features between the target subject and the direct object reflexive across the three experiments of their study (Experiment 1: animacy + gender; Experiment 2: animacy + number; Experiment 3: gender + number). Results from these experiments showed that anaphoric reflexives are selectively prone to attraction/interference, depending on the strength of the feature match between the antecedent and the anaphoric reflexive. Strong interference was observed to occur when the licensor and the reflexive mismatched in multiple features. Parker

and Phillips (2017) further suggested a cue-based memory retrieval mechanism that uses a cue-combinatorics scheme that gives stronger preference to syntactic cues than morphological cues as in the case of anaphoric reflexive resolution.

Moreover, failure to find interference effects in reflexive-antecedent dependencies in some of the earlier studies (e.g., Dillon et al., 2013; Sturt, 2003) was argued to be caused by the absence of a corresponding syntactic role between the distractor and the target antecedent. Patil et al. (2016) found that interference effects arise when the target and the distractor are assigned similar syntactic roles (e.g., subject role). Findings from the studies by Patil et al. (2016) and Parker and Phillips (2017) suggested that reflexives, though selective, are also affected by interference effects. The findings provide further evidence to the validity of a cue-based retrieval mechanism that allows deployment of both structural and non-structural cues during the resolution of dependencies, albeit preferences for the type of cues vary across different types of syntactic configurations.

As stated earlier, the cue-based memory retrieval mechanism was extended from the ACT-R framework. ACT-R theory has been applied in numerous studies on education (e.g., learning behavior), cognitive development, language comprehension and production, etc. Lewis and Vasishth (2005), Badecker and Lewis (2007), and more recently Engelmann et al. (2019) developed cue-based memory retrieval models from the ACT-R theory with a focus on agreement computation and dependency resolution. Under a cue-based memory retrieval mechanism, memory-encoded items have different activation rates depending on the degrees of feature match and mismatch between the items and the retrieval cues. Items with higher cue-matching scores are associated with higher levels of activation, lower chance of memory decay, and thus higher probability of retrieval. This cue combination architecture adopts a linear

method of cue combination (Parker, 2019). Following a linear cue combination method, if an item has a great number of matching cues, the item will have higher match scores, thus higher summation of total cue strength.

Models that apply a linear cue combination method include the original LV05 model suggested by Lewis and Vasishth (2005). By contrast, another approach to cue strength combination follows a non-linear method which does not simply view total cue strength as a linear summation of each associative cue, but more as a "super-additive/exponential growth" of matching cues (Parker, 2019, p. 3). The original LV05 model of the memory access mechanism (Lewis & Vasishth, 2005) has been challenged for its simplistic cue combination architecture. Parker (2019), for that reason, advocated the development of processing models that feature both a cue-combination and a cue-integration architecture. A large proportion of early memory retrieval models embraced either a linear or nonlinear method of combining cues (Parker, 2019), which has been recently reported to inevitably fall short on giving a complete memory access architecture that could illuminate the types of cues deployed and prioritized by the parser during retrieval.

The two methods (i.e., linear and nonlinear) of cue combination differ in how cues are weighted. Equation (1) illustrates the strength association that each cue contributes to the activation of an item following the linear method (Parker, 2019).

$$A_i = \sum_{j=1}^{n} W_j S(Q_j, I_i) \qquad (1)$$

As shown in Equation (1), A_i is the activation of item I_i, which is calculated as the direct summation of associative strength (S) between the features of item I_i and the retrieval cue Q_j, while W_j is the weight given to the cue. The equation depicts an additive function, which

represents the escalated activation by virtue of a linear combination, or in other words, an independent contribution of cues.

On the contrary, the nonlinear method takes into account the multiplication rather than the summation of cue strengths as specified in Equation (2).

$$A_i = \prod_{j=1}^{n} S(Q_j, I_i)^{W_j} \qquad (2)$$

Models that adopt a nonlinear cue integration method do not treat the combination of cues as an independent contribution of individual cues, but instead, as concurrence of interdependent cues. In this regard, an item that matches all of the cues would potentially outweigh the summation of all partial matches, and thus more likely to be activated, and then retrieved (Parker, 2019).

Due to its oversimplification of memory access architecture which failed to capture the interference effects and processing phenomena that went beyond its own capabilities, the original LV05 model (Lewis & Vasishth, 2005) was evaluated and upgraded into the LV05 + IP (i.e., item prominence) + MAC (i.e., multi-associative cues) model (Engelmann et al., 2019) to better unveil the cognitive processes underlying sentence comprehension. The original LV05 was developed on the ground of the ACT-R architecture which relied on a number of assumptions that, over the last decade, has failed to explain some empirical phenomena in a number of past and recent experiments. The original assumptions are as follows (Engelmann et al., 2019, p. 14).

1. The base-level activation of items in memory is a function only of decay or reactivation through study-relevant retrieval events. Other influences are usually ignored.

2. The fan effect (the inhibitory interference effect caused by cue overload) is a function of the number of items that match a specific retrieval cue, independent of their activation.

3. The associative strength between a retrieval cue and a memory item is based on a binary (match/mismatch) one-to-one mapping between the cue and a feature value.

To better account for the underlying aspects of cognition, the extension of the LV05 revised the oversimplifying assumptions mentioned above in a more relaxing fashion. Below are the assumptions embedded in the updated LV05 model (Engelmann et al., 2019, p. 14):

1. The base-level activation of items in memory (i.e., accessibility) is affected by – in addition to recency – their prominence in the current context, that is, their relevance/salience in terms of syntactic relations in a sentence or information-structural and discourse properties.

2. The strength of any interference effect – including the fan effect – is not simply a function of the presence versus the absence of a distractor, but changes as a function of the distractor's activation in memory relative to the target.

3. The associative strength between a retrieval cue and a memory item can be the result of multiple cues being associated with multiple features at variable degrees. Cue-feature associations are based on associative learning through language experience.

The revised LV05 model was extended on the account of meta-analysis of past research (see Jager, Engelmann, & Vasishth, 2017), thus its application could converge on the processing of various syntactic constructions. This section, however, narrows down the application of the updated LV05 to the research focus of the study only – which is resolution of reflexive-antecedent dependencies.

Comparison between the original and the revised LV05 was conducted using examples of reflexive constructions taken from the study by Sturt (2003).

(33) a. *Target-match; distractor-mismatch (no interference)*

The surgeon$^{+\text{MASC}}_{+\text{CCOM}}$ who treated Jennifer$^{-\text{MASC}}_{-\text{CCOM}}$ had pricked himself$\{^{\text{MASC}}_{\text{CCOM}}\}$…

b. *Target-match; distractor-match (interference)*

The surgeon$^{+\text{MASC}}_{+\text{CCOM}}$ who treated Jonathan$^{+\text{MASC}}_{-\text{CCOM}}$ had pricked himself$\{^{\text{MASC}}_{\text{CCOM}}\}$…

c. *Target-mismatch; distractor-mismatch (no interference)*

The surgeon$^{-\text{FEM}}_{+\text{CCOM}}$ who treated Jonathan$^{-\text{FEM}}_{-\text{CCOM}}$ had pricked herself$\{^{\text{FEM}}_{\text{CCOM}}\}$…

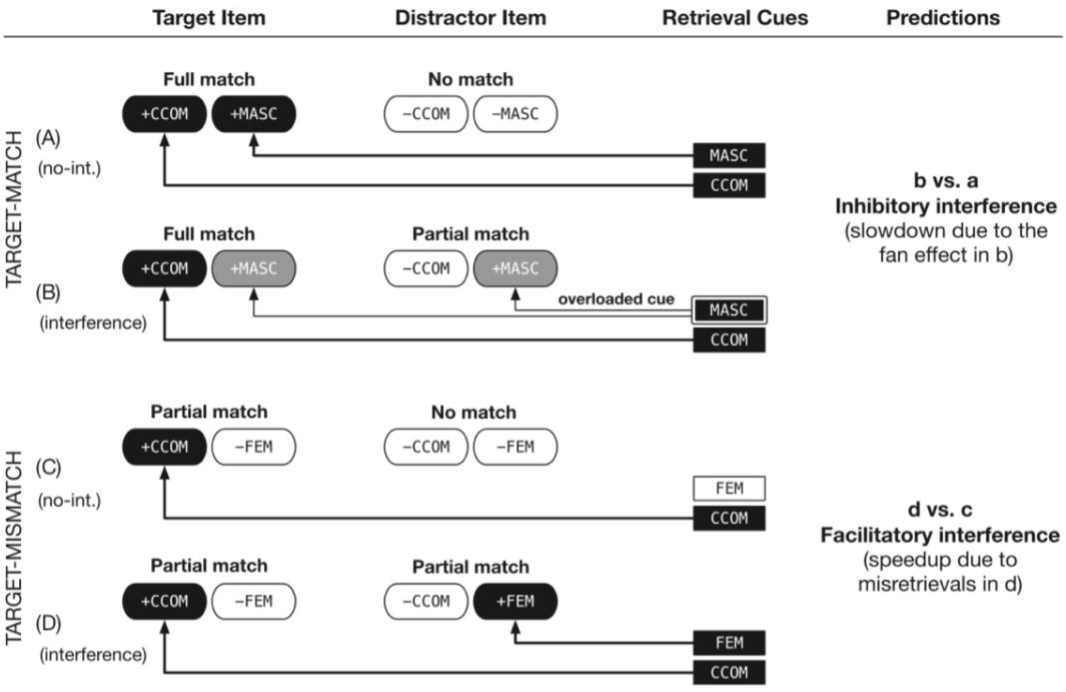
d. *Target-mismatch; distractor-match (interference)*

The surgeon$^{-\text{FEM}}_{+\text{CCOM}}$ who treated Jennifer$^{+\text{FEM}}_{-\text{CCOM}}$ had pricked herself$\{^{\text{FEM}}_{\text{CCOM}}\}$…

The original LV05 classified interference effects into two types: *inhibitory interference* (i.e., slow down) and *facilitatory interference* (speed up) which were expected to manifest in the reading times of the experimental items. Figure 1 (Engelmann et al., 2019, p. 8) graphically illustrates the predictions of the original LV05 for the examples given in (33). The darker shapes indicate the better cue match and the stronger degree of item activation, while the amount of spreading activation (which will be discussed below) is represented by the thickness of the arrowed lines.

Figure 1

Spreading Activation According to ACT-R/LV05 (Engelmann et al., 2019, p. 8)



Following the predictions of the original LV05, all items encoded in memory compete for retrieval at the same time, and the "winner" is the item that has the strongest degree of activation. The original LV05 suggested that each item *i* has its own base-level activation (see Equation 3) as well as the added spreading activation (Equation 4) which is shared by all items that feature-match the cue(s) at the time of retrieval.

$$B_i = \ln\left(\sum_{j=1}^{n} t_j^{-d}\right) + \beta_i \qquad (3)$$

In Equation 3, B_i is the base-line activation of item *i*; t_j is the past time since the j^{th} activation; *d* is the decay rate ($d = 0.5$ in ACT-R); β_i represents the resting state of activation of item *i; n* is the number of times that item *i* was previously retrieved.

Under the original LV05 model, all cues are equally weighted, which makes the cue's weight W_j the result (difference) of the subtraction $\frac{1}{number\ of\ cues}$. Following Equation 4, the spreading activation S_i of the item *i* is the sum of all matching cues *j* taking into consideration the associative strength S_{ji} between item *i* and cue *j* given the cue's weight W_j.

$$S_i = \sum_j W_j S_{ji} \qquad (4)$$

When multiple items, including the target and the distractor(s), feature-match the cue(s) as in the interference configurations in example (33), cue overload will arise, leading to reduction in the associative strength between the item and the cue (S_{ji}). *MAS* is the maximum associative strength between the cue and the item as shown in Equation (5). The number of items sharing the same features with the cue is called a *fan*. *Fan effect* (Anderson, 1974, as cited in Engelmann et al., 2019, p. 9) is the term used to refer to the reduction of activation to the target due to cue overload.

$$S_{ji} = MAS - \ln(fan_j) \qquad (5)$$

Compared to no-interference constructions in examples (33a) and (33c), the fan effect in the target-match interference configuration (example 33b) caused reduction in the target's activation, which then consequently leads to an increase in retrieval latency of item *i* (RT_i), or in other words, inhibitory interference. RT_i is represented by the negative exponential function with *F* is the latency factor and *f* is the latency exponent (Equation 6).

$$RT_i = Fe^{-(f \times A_i)} \qquad (6)$$

Depending on the extent to which the target and the distractor differ in activation strength, there is a possibility of misretrieval as a result of a fan effect. Misretrieval is more

likely to occur in the interference configuration of the target-match condition in example (33). By contrast, the original LV05 did not predict a reduction in the activation of the target in the interference configuration of the target-mismatch condition. Explanation given to the absence of inhibitory interference in example (33d) is that the target and the distractor – each only matches one of the two cues (i.e., FEM and CCOM), thus the two items relatively receive the same degree of activation and the equal probability (0.5) of being retrieved. The original LV05 further predicts a speed-up (i.e., facilitatory interference) in retrieval latency in example (33d) compared to (33c) across multiple trials.

However, such predictions of the original LV05 (i.e., inhibitory interference in target-match condition, facilitatory interference in target-mismatch condition) are not always in line with findings from other research, which motivated the revision of the LV05 (Elgemann et al., 2019). Taking into account item prominence (IP) and multi-associative cues (MAC), the LV05+IP+MAC model managed to capture the inhibitory effects in the target-mismatch configuration in reflexive dependencies (see Jager et al., 2017).

By item prominence (with a focus on discourse status and syntactic position), the revised LV05 suggested that items with high prominence are encoded longer in memory, thus having shorter latency and higher probability of retrieval than items which are less prominent. The inclusion of item prominence in the extended model can capture both the inhibitory and the facilitatory effects in the target-match condition while the original model could only predict the former one.

Instead of treating a match between a cue and the feature of an item as a binary, one-to-one categorical relation as suggested by the original LV05, the updated model takes into consideration the possibility that a cue can match multiple features at varying degrees

(Engelmann et al., 2019). Therefore, the associative strength S_{ji} between item *i* and cue *j* now accounts for the probability *P(i|j)* of item *i* being considered for retrieval as shown in Equation 7.

$$S_{ji} = MAS + \ln[P(i|j)] \quad (7)$$

Equation 8 illustrates the probability *P(i|j)* as the match quality Q_{ji} of item *i* with cue *j* subtracted by the total match quality Q_{jv} of all items *v* with cue *j*.

$$P(i|j) = \frac{Q_{ji}}{\sum_{v \in Items} Q_{jv}} \quad (8)$$

The match quality Q_{ji} of item *i* with cue *j* is determined by the associative strength of cue *j* and all features K_i of item *i*:

$$Q_{ji} = \sum_{k \in K_i} M_{jk} \quad (9)$$

Engelmann et al. (2019) stated that a fan effect is predicted as long as there is/are shared feature(s) between the item(s) and the cue, even in cases where the items do not have any features in common. Under the circumstance where item *i* has only feature *f* that matches cue *q*, the match quality $Q_{q,i}$ gets the value of 1, $Q_{q,i'} = 0$, thus $\sum_{v \in items} Q_{q,v} = Q_{q,i} + Q_{q,i'} = 1 + 0 = 1$; item *i'* has only feature *f'* that matches cue *q'*, then $Q_{q',i'} = 1$, $Q_{q',i} = 0$, $\sum_{v \in items} Q_{q',v'} = Q_{q',i'} + Q_{q'i} = 1 + 0 = 1$; making $P(i|j) = P(i|q) = \frac{1}{1} = 1$. Plugging the values into the above equations, the spreading activation $S_{q,i}$ from cue *q* to item *i* equals *MAS* which is the maximal associative strength between the cue *q* and the item *i*: $S_{q,i} = MAS + \ln[P(i|q)] = MAS + \ln[1] = MAS$.

In another situation in which item *i* receives activation from not only cue *q*, but also cue *q'*, the associative strength $S_{q,i}$ and $S_{q',i}$ of item *i* will be calculated respectively as follows: MAS – 0.41; MAS – 1.1 (Engelmann et al., 2019, p. 27). Details of how $S_{q,i}$ and $S_{q',i}$ of item *i* are

calculated are discussed in depth in section 5.4.1 with respect to the experimental items of the current study. As an item gets activation from multiple cues, and a cue is also distributed to match the corresponding feature of multiple items, the spreading activation $S_{j,i}$ of an item *i* given cue *j* no longer matches up the maximal associate strength *MAS*.

The updated LV05 added the component p_i (i.e., prominence of item *i*) to the item's base-level activation B_i, which was depicted in Equation 10.

$$B_i = \ln\left(\sum_{j=1}^{n} t_j^{-d}\right) + \beta_i + p_i \qquad (10)$$

The addition of the component p_i indicates the increased strength of activation that item *i* receives. A prominent item *i* has strong interference with other items during retrieval. The more prominent an item is, the higher the activation boost, and the more likelihood of retrieval.

Compared with the earlier version, the revised LV05 can account for the inhibitory and facilitatory effects in both target-match and target-mismatch conditions over different types of syntactic constructions (Table 1). By including the principles of item prominence and multi-associative cues, the new LV05 model could predict a broad range of experimental outcomes from empirical research as well as from future work.

Table 1

Retrieval Models with Different Methods of Cue Combination

	LV05 (Lewis & Vasishth, 2005)	**LV05 + IP + MAC (Engelmann et al., 2019)**	**Others (Parker, 2019)**
Method of cue combination	Linear (summation of cue strengths)	Linear (summation of cue strengths)	Non-linear (multiplication of cue strengths)
Assumptions of retrieval process	- The base-level activation of items in memory is a function only of decay or reactivation through study-relevant retrieval events. - The fan effect (the inhibitory interference effect caused by cue overload) is a function of the number of items that match a specific retrieval cue, independent of their activation. - The associative strength between a retrieval cue and a memory item is based on a binary (match/mismatch) one-to-one mapping between the cue and a feature value. (Engelmann et al., 2019, p. 14)	- The base-level activation of items in memory (i.e., accessibility) is affected by – in addition to recency – their prominence in the current context (i.e., their relevance/salience in terms of syntactic relations in a sentence or information-structural and discourse properties). -The strength of any interference effect – including the fan effect – is not simply a function of the presence versus the absence of a distractor, but changes as a function of the distractor's activation in memory relative to the target. -The associative strength between a retrieval cue and a memory item can be the result of multiple cues being associated with multiple features at variable degrees.	-Retrieval: sensitive to conjunctions of cues, rather than occurrence of individual cues → target items that match all of the cues are favored exponentially more than partially matching target items. (Parker, 2019, p. 5) -Cue strengths are multiplied, rather than summed, which causes a much greater reduction in the activation for partial matches that occurs with a linear scheme → interference from nontarget partial matches less likely. (Parker, 2019, p. 8)

2.1.3. Resolution of Reflexive-Antecedent Dependencies

Studies on resolution of reflexive-antecedent dependencies have taken into consideration all the theoretical accounts of anaphora resolution that were discussed and reviewed in the preceding sections: Binding theory (e.g., Badecker & Straub, 2002; Sturt, 2002; Cunnings & Sturt, 2014), agreement (e.g., Kreiner et al., 2008), discourse prominence (e.g., Garnham, 2001; Garrod & Sanford, 1994; Nicol & Swinney, 2003, and memory access (e.g., Cunnings & Felser, 2013; Dillon, 2011; Patil, et al., 2016; Parker & Phillips, 2017).

As reviewed in section 2.1.2.1, the processing and resolution of anaphoric reflexives were argued to be primarily governed by syntactic constraints, specifically Binding Principle A (Chomsky, 1981, 1986), c-command (Reinhart, 1976, 1983) over morphological constraints (e.g., Dillon, 2011). In addition to structural cues, comprehenders also make use of non-structural cues (e.g., gender, animacy, etc.) to access the target licensor (Patil et al., 2016; Parker & Phillips, 2017). However, it cannot be inferred from the simultaneous presence of multiple constraints that their effects on retrieval are also simultaneous (Sturt, 2003). Under a cue-based memory retrieval mechanism (Lewis & Vasishth, 2005; Engelmann et al., 2019), non-structural cues may accelerate retrieval times, and in some circumstances may result in misretrieval of inaccessible antecedents (Patil et al., 2016).

Nicol and Swinney (1989) suggested a number of hypotheses concerning the time course of binding constraints. The *binding-as-initial-filter* hypothesis assumes that binding principles are applied at the very early stage of processing and continue to influence all the later stages (Nicol & Swinney, 1989; as cited in Sturt, 2003, p. 543). *Binding-as-late-filter* suggests that Binding theory is still applied, but at the later processing stages, functioning as a *filter* to resolve the misretrieval of an inaccessible antecedent. Misretrieval, in this case, is often caused by the

misinterpretation of lexical information during early processing when syntactic information governing the application of Binding theory is not yet readily available. *Binding-as-defeasible-filter*, on the other hand, indicates that Binding theory is applied during early processing; however, in the later stages, comprehenders still have access to inaccessible antecedents as a result of cue overload or distractor prominence. Misretrieval of antecedents was also attributed to the multiple constraints account (e.g., Spivey & Tanenhaus, 1998; as cited in Sturt, 2003, p. 543) which hypothesizes that all relevant constraints are deployed simultaneously during processing to guide the licensing of the target antecedent.

(34) (***Jonathan/Jennifer***) was pretty worried at the City Hospital. (***He/She***) remembered that the surgeon had pricked (***himself/herself***) with a used syringe needle. There should be an investigation soon.

(35) (***Jonathan/Jennifer***) was pretty worried at the City Hospital. The surgeon who treated (***Jonathan/Jennifer***) had pricked (***himself/herself***) with a used syringe needle. There should be an investigation soon.

Sturt (2003) investigated the time course of binding constraints in the processing of anaphoric reflexives using experimental stimuli as shown in examples (34) and (35). The study also examined the effects of violations of the binding principles on comprehenders' final interpretation of the experimental sentences. Findings from the experiments showed a strong interference effect: when the inaccessible antecedent matched the reflexive in gender, comprehenders were more likely to misinterpret the experimental sentences. The interference effect was observed to be even more robust when the target antecedent mismatched the gender of the reflexive. Conclusions from Sturt's (2003) experiments demonstrated that binding constraints were applied at early processing stages, but later were violated by the interference of

a discourse-focused inaccessible antecedent. The finding provided additional support for the *binding-as-defeasible-filter* hypothesis.

Other studies also generated similar findings regarding the early/late application of Binding theory and the interference effects from distractors that match the target antecedents in gender features (e.g., Badecker & Straub, 2002; Cunnings & Felser, 2013). Studies that found no attraction effects in resolution of anaphoric reflexives include Dillon et al.'s (2013); Xiang et al.'s (2009); etc. as reviewed in section 2.1.2.3. Dillon et al. (2013) used experimental sentences of which syntactic structures resemble those used in Sturt's (2003) research, yet the researchers found no effects of inaccessible antecedents in their eye-tracking experiments. Xiang and colleagues (2009) observed an elicitation of a P600 among comprehenders immediately upon their encounter with the anaphoric reflexive that matched the accessible antecedent in gender. Furthermore, such a peak in comprehenders' electrical brain activity was not weakened by the gender match between the reflexive and the syntactically inaccessible antecedent, which supported the proposal that only structural cues are deployed during retrieval, making other non-structural cues (e.g., gender feature) trivial in the search for the target antecedent.

Patil et al. (2016) critically analyzed the earlier studies that found no interference from non-structural constraints during retrieval, and pointed out that the absence of attraction effects in those studies (e.g., Dillon et al., 2013; Sturt, 2003) might be due to the different syntactic roles assigned to the distractor and the target antecedent. Patil et al. (2016) used experimental stimuli which were modified from Sturt's (2003) to increase the strength and probability of attraction effects. The gender of the distractor and the gender of the reflexive were manipulated across the four conditions of the experimental items (36a, b, c, d). In addition to that, Patil et al.

(2016) also increased the structural similarity between the accessible and inaccessible antecedents by assigning them a subject role in the sentence.

(36) (a) The ***head engineer*** that ***Peter*** had visited in the factory convinced ***himself*** that the building was safe.

(b) The ***head engineer*** that ***Nancy*** had visited in the factory convinced ***himself*** that the building was safe.

(c) The ***head engineer*** that ***Nancy*** had visited in the factory convinced ***herself*** that the building was safe.

(d) The ***head engineer*** that ***Peter*** had visited in the factory convinced ***herself*** that the building was safe.

Parker and Phillips (2017), as reviewed in section 2.1.2.3, also provided counter-evidence to the previous studies that found no interference effects in reflexive-antecedent dependencies (e.g., Cunnings & Sturt, 2014; Dillon et al., 2013; Sturt, 2003; Xiang et al., 2009). Parker and Phillips (2017) concluded that "reflexives are indeed susceptible to attraction" (p. 284), albeit selectively, and that the strength of attraction effects depends on the degree of feature match between the reflexive and the target antecedent.

Results from Patil et al.'s (2016) and Parker and Phillips's (2017) studies are consistent with the prediction of the cue-based retrieval model, such that non-structural cues are deployed in parallel with structural cues during retrieval. Mixed findings from empirical research concerning the deployment of different types of cues during retrieval motivate the need for further research on the application of a memory access mechanism focusing on constraints and selection of cues that guide referential resolution in real time sentence processing.

2.2. Prosody and Processing

Prosody is a collective term that covers a wide range of suprasegmental acoustic features occurring in human fluent speech including syllable duration, word stress, pitch, intonation, and timing variations (Hoyte, Brownell, & Wingfield, 2009). The present study focuses on timing variations, also understood as prosodic boundaries (pauses) between different phrasal segments of a sentence. Prosodic boundaries are predicted to facilitate comprehenders' syntactic analysis of the sentence since prosodic boundaries occurring in natural speech, to a certain extent, correspond with the syntactic structure of the sentence (Cutler, Dahan, & Van Donselaar, 1997; as cited in Pratt, 2015, p. 38; Hoyte et al., 2009). Prosodic contours might also have certain effects on discourse planning and production in simultaneous speech (Kraljic & Brennan, 2005). Different or inappropriate use of prosody, to a certain extent, affects intelligibility in communication since prosodic cues are used as indicators of meaning in discourse (Kang, Johnson, & Kermad, 2021). For that reason, manipulating prosodic contours may interfere with the processing of a sentence, and sometimes resulting in sentence misinterpretation. One of the aims of the current study is to get insight into how manipulations of the implicit prosodic contour of a sentence during the silent reading affect comprehenders' processing and interpretation of the sentence. It was expected that different prosodic patterns projected onto the experimental stimulus would either facilitate or hinder processing, and that this would be reflected in participants' responses in a grammaticality judgment task and in a comprehension probe task.

A large body of research has investigated prosody in comprehension, focusing on how comprehenders process and utilize prosodic information in sentence processing given the assumption that prosodic grouping reflects the hierarchical syntactic relations between different

discourse segments in a given speech (Cutler, Dahan, & van Donselaar, 1997; Gee & Grosjean, 1983; Nespor, Shukla, van de Vijver, Avesani, Schraudolf, & Donati, 2008). Researchers supporting this account view prosodic boundaries as cues that guide processing and comprehension. As for production, changes in the prosodic grouping of an utterance reflect the changes in the speaker's intention in the production of an utterance. An alternative account, in contrast, suggests that there is only one plausible prosodic contour that is dictated by the syntactic structure of the sentence regardless of the context in which the sentence is used (Fraizer, Carlson, & Clifton Jr., 2006). Much research has been conducted to provide further evidence for the first account, justifying the interplay between prosody and syntax (e.g., Cutler et al., 1997; Matsui, Nakamura, Utsumi, Sasaki, Koike, Yoshidi, Harada, Tanabe, & Sadato, 2016; Lieberman, 1967; Schafer, 1997).

(37) (a) [The bus driver angered the rider] [with a mean look].

(b) [The bus driver angered] [the rider with a mean look].

Schafer (1997) observed that prosodic breaks determine the attachment of the prepositional phrase *with a mean look* in (37) to either the noun phrase *the rider* (low attachment reading) or *the bus driver* (high attachment).

(38) The friend of the girl that sits at the cafe is talkative.

With respect to ambiguous relative clauses as (38), research on syntactic attachment preferences by English native speakers (L1ers) and nonnative speakers (L2ers) found that L1ers are more likely to gravitate towards low attachment readings (*the girl*) while L2ers have preferences for high attachment (*the friend*) (e.g., Dinctopal-Deniz, 2010). However, comprehenders may be biased towards certain prosodic groupings, and in certain cases,

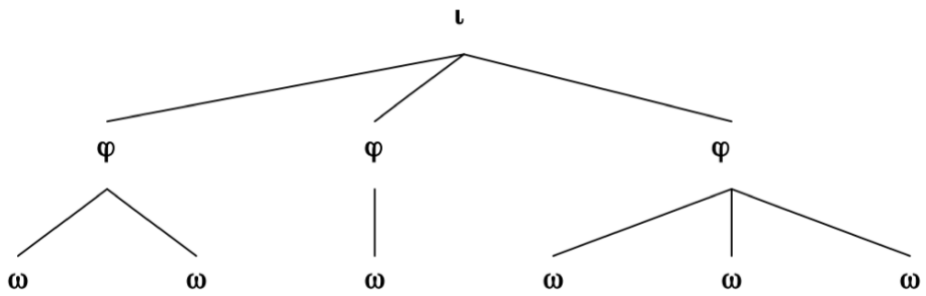
comprehenders' final decision may differ from their initial attachment preference (Fernández, 2007; Maynell, 1999, 2000).

A large body of research found that prosody facilitates the syntactic interpretation of an utterance. Association between prosodic representations and syntactic structure was based on the theory of Hierarchical Intonation Structure (e.g., Schafer, 1997; Speer, Shih, & Slowiaczek, 1989; as cited in Pratt, 2015, p. 39). Holding an opposing view, the Strict Layer hypothesis (Hayes, 1989; Inkelas, 1990; Nespor & Vogel, 1986; Pierrehumbert & Beckman, 1988; Selkirk, 1981, 1984, 1995) did not support the relation between prosody and syntax. In other words, the hypothesis is strictly restricted to phonological relations, such that "no inherent relation is assumed to exist between the prosodic category types found in phonological representations and the category types of syntactic representation" (Selkirk, 2011, p. 3).

Under the Strict Layer hypothesis, the prosodic hierarchy of an utterance (U) contains the following categories: intonational phrase (ι); phonological phrase (ϕ); prosodic word (ω); foot; syllable (Figure 2).

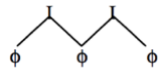
Figure 2

Prosodic Hierarchy of an Utterance (Hayes, 1989; Inkelas, 1990; Nespor & Vogel, 1986; Pierrehumbert & Beckman, 1988; Selkirk, 1981, 1984, 1995, 2011)



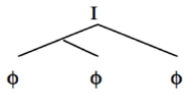
The hypothesis was formulated to primarily demonstrate the relations between the constituents of different prosodic categories in a prosodic structure (Selkirk, 2011). Selkirk (1981) initially proposed the Strict Layer hypothesis stating that "a constituent of category-level *n* in the prosodic hierarchy immediately dominates only (a sequence of) constituents at category-level *n-1* in the hierarchy" (as cited in Selkirk, 2011, p. 3).

Though the Strict Layer hypothesis was well received in the literature, opponents to the hypothesis (e.g., Ladd, 1984, 2008; Tokisaki, 2001) have argued that the hypothesis did not take into consideration a number of other possibilities of prosodic hierarchy, specifically as follows:

- Multiple domination: 

- Heterogeneous sisters: 

- Skipping of levels: 

- Unlabeled nodes: 

- Recursion: 

Researchers holding alternative views to the Strict Layer hypothesis raised a number of questions concerning the nature, the types, and the number of prosodic categories (e.g., Nespor & Vogel, 1986) within an utterance (Tobisaki, 2001). Though further research is needed, the interplay between syntax and prosody has been supported by findings from earlier research (e.g., Fraizer et al., 2006; Heffner & Slevc, 2015; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Selkirk, 2011). Selkirk (2011) argued for a language-dependent ranking regarding the matching degree between the prosodic constituent structure of a sentence and the structure of its syntactic constituents. The interplay between prosody and syntax will be discussed in more depth in section 2.2.2.

2.2.1. Prosody in Comprehension

Information conveyed through the prosodic structure of an utterance includes word stress, discourse focus, intention of the speaker, etc. (Ladd, 2008). As reviewed in the previous section, the prosodic grouping of a sentence has certain effects on sentence comprehension, which can be attributed to the alignment between the prosodic parse and the syntactic structure of the sentence. The prominence of a word in a phrase or a sentence is represented by either high (H*) or low (L*) pitch accents which are the primary units of prosody. Word groups, on the other hand, align with prosodic phrases. The prosodic contour of a sentence, as a result, plays an important role in the comprehension of the sentence, particularly in cases of syntactic ambiguity (e.g., Kjelgaard & Speer, 1999; Kraljic & Brennan, 2005). Additionally, prosodic parsing is also claimed to support the retention of information in memory during sentence processing (Kjelgaard & Speer, 1999; Kreiner, 2006).

It is still a matter of open debate among linguists as to where and when prosody should be used. Prosodic features are language-dependent. For a language such as English, most of the content words of a sentence receive prominent stress on particular syllables, which makes word stress one of the prosodic aspects that is mandatory in English. Apart from word stress, prosodic breaks of an utterance are also obligatory (e.g., *Before we left the class,...*) (Frazier et al., 2006).

Studies on infants' language acquisition have shown that infants are sensitive to prosodic information (Jusczyk, Hirsh-Pasek, Nelson, Kennedy, Woodward, & Piwoz, 1992) and that children predominantly make use of prosodic information to parse natural speech into phrases and words to identify the important syntactic units of an utterance (Morgan & Saffran, 1995; Jusczyk, 2003). Young infants are able to utilize prosodic information in connected speech to guide their syntactic interpretation of the sentence (Gleitman & Wanner, 1982), and even

"encode information from the speech signal into memory" (Mandel, Jusczyk, & Nelson, 1994; as cited in Pratt, 2015, p. 42).

Adults also exploit prosodic information in auditory sentence comprehension, and their ability to recognize words relies on the salient syllables of the words used in the discourse (Ladd, 1984). Word stress was predicted to affect response accuracy in experimental tasks. Studies using experimental stimuli that manipulated word stress patterns (i.e., *incorrect* versus *correct* word stress) found that higher accuracy was associated with the use of correct word stress (e.g., Slowiaczek, 1990). The acceptability rating of a word in a given context, nevertheless, was found to be independent from the stress pattern of the word. Provided that the word is semantically suitable for the given context, the stress pattern of the word does not seem to play any significant role in guiding acceptability judgments (Slowiaczek, 1990). However, effects of stress patterns were suggested to be language-dependent. Speakers of languages other than English (e.g., German), in contrast, were observed to be susceptible to manipulations of word stress in processing (e.g., Domahs, Wiese, Bornkessel-Schlesewsky, & Schlesewsky, 2008).

Frazier et al. (2006) introduced the Optional Prosodic Boundaries hypothesis. The hypothesis predicts that comprehenders will have different preferences for syntactic attachments of a syntactically ambiguous sentence. Frazier et al. (2006) also argued that prosodic grouping was partly dictated by the syntactic structure and the length of the phrasal unit. Prosody helps comprehenders retain information in memory during processing, especially in auditory processing. Parsing information into a string of digits or a chunk of words makes it easier for the comprehender to hold information in memory, thus aiding the processing of the sentence as a whole. Further evidence regarding this observation was also found in nursery rhymes for young

infants, and in research on the effects of pitch accent on attachment preferences (Schafer, Carter, Clifton, & Frazier, 1996). Those observations support the proposal that comprehenders make use of prosodic information during parsing to guide their decisions in constructing an interpretation of an utterance.

2.2.2. Interplay between Syntax and Prosody

It has been justified in the literature that the prosodic grouping of a sentence, to a certain extent, corresponds with its syntactic parsing (Gee & Grosjean, 1983; Ferreira, 1988; Selkirk, 1984; Snedeker & Trueswell, 2003; Watson & Gibson, 2004). Prosodic information has been shown to play a crucial role in disambiguating syntactically ambiguous sentences of which constructions include the three most common types: noun phrase (NP) ambiguity, prepositional phrase (PP) ambiguity, and relative clauses.

2.2.2.1. NP-ambiguity

A relatively large body of research has focused on the correlation between prosodic parsing and syntactic boundaries in the comprehension and production of NP-ambiguity sentences (e.g., Allbritton, Mckoon, & Ratcliff, 1996; Lehiste, 1973; Scott, 1982). Results from those studies demonstrated that in sentence production, participants varied in duration of words and/or duration of pauses to indicate different preferences for NP attachments. Prosodic groupings were predicted to guide participants in the disambiguation of a syntactically ambiguous sentence (Beach, Katz, & Skowronski, 1996; Price et al., 1991; Scott, 1982).

(39) (a) For our parties, we invite [David and Pat] or [Bob], but not all three.

(b) For our parties, we invite [David] and [Pat or Bob], but not all three.

Allbritton et al. (1996) expected to find phrase-final lengthening that determines the attachment of the NP (i.e., lengthening of the critical NP *Pat* in (39a) as compared to *Pat* in

(39b)). Participants (including both trained and untrained speakers) in Allbritton et al.'s study did not exhibit consistency in the prosodic parsing of the experimental items. However, the prosodic cues participants used did facilitate the processing and resolution of the given ambiguous sentences.

Beach, Katz, and Skowronski (1996) reported similar findings, indicating participants' use of prosodic information as cues that guide the interpretation of the experimental sentences. The stimuli used in their experiments differ in prosodic groupings (e.g., *[pink and green] and [white]* vs. *[pink] and [green and white]*). Participants of the study were both adults and children (the children were at the age of five to seven). Results showed that children varied in their pitch and duration of pauses just as adults did in the disambiguation of the stimuli. Contrary to Beach et al.'s findings, Katz, Beach, Jenouri, and Verma (1996) found that the five year olds and seven year olds in their study did not perform as well as the adults did in the use of prosodic cues as required by the experimental task. Nevertheless, such difference in children's decisions on prosodic parsing might not be attributed entirely to age because children were observed to differ even within and across age groups (five to 14) in other studies (e.g., Wells, Peppe & Goulandris, 2004).

2.2.2.2. PP-ambiguity

Example (37) demonstrates a typical example of PP-attachment ambiguity. For PP ambiguity, Watson and Gibson (2004) proposed the hypothesis of Anti-attachment which was expected to eliminate the negative effects of prosodic boundaries on sentence processing. Under this theory, the prosodic boundary inserted right after the critical phrase of the sentence (e.g., *the paintings* as in (40a) and *the paintings of the landscape* as in (40b)) was expected to prevent further attachments to the PP segment.

(40) (a) [An artist arranged a donation of the paintings] [of the landscape] [to the museum].

(b) [An artist arranged a donation of the paintings of the landscape] [to the museum].

Watson, Breen and Gibson (2006) found that comprehenders still benefited from the insertion of a prosodic break between the two PPs (e.g., *of the landscape* and *to the museum*) during processing. When prosodic breaks were inserted into a sentence, participants found the sentence more comprehensible compared to when there were no prosodic breaks.

2.2.2.3. Relative clauses (RCs)

Though prosodic boundaries were predicted to facilitate comprehenders during processing, there exist open questions concerning the extent to which comprehenders use prosodic information in the resolution of syntactic ambiguities. While the Anti-attachment hypothesis (Watson & Gibson, 2004; Watson, Breen & Gibson, 2006) is in favor of instant deployment of prosodic information in processing, the Informative Boundary hypothesis (Clifton, Carlson, & Frazier, 2002), on the contrary, does not treat the interpretation of a prosodic break simply within its absolute boundary, but instead with consideration given to other prosodic boundaries occurring in the same sentence. In addition to studies using NP and PP-attachments, relative clause (RC) attachments have also been investigated to offer insights into the interplay between prosody and syntax in the disambiguation of syntactically ambiguous sentences.

(41) Someone shot [the servant]$_{NP1}$ of [the actress]$_{NP2}$ [who was on the balcony].

(a) The servant was on the balcony.

(b) The actress was on the balcony.

Example (41) has two possible readings depending on the attachment of the RC to either the first NP (i.e., *high attachment – HA*) or the second NP (i.e., *low attachment – LA*). Attachment preferences were observed to vary cross-linguistically: English speakers are more likely to have preferences for LA (Fernández, 2003) according to the Late Closure Principle (Frazier, 1979), whereas Spanish speakers prefer HA to LA interpretation, which was reflected in the increased reading times in a forced LA interpretation (Cuetos & Mitchell, 1988; as cited in Biau, Fromont, & Soto-Faraco, 2018, p. 2).

In addition to RC attachment preferences, the interpretation of RC ambiguities was argued to also rely on other linguistic factors such as the lexical-semantic and pragmatic information of the sentence (Fernández & Sekerina, 2015). For example (41), responses to the comprehension question *Who was on the balcony?* may also be affected by participants' interpretation of the lexical-semantic aspect of the two NPs *the servant* and *the actress,* which may consequently guide participants' decision on who was more likely to be on the balcony – *the servant* or *the actress*? Furthermore, when prosodic parsing was held constant across experimental stimuli, variations in the length of the RC were observed to have an effect on attachment preferences: comprehenders had a tendency to bias towards HA preference in cases of long RCs, and LA preference for short RCs (Fernández & Sekerina, 2015). Biau et al. (2018) provided further evidence of a syntax-prosody mapping in syntactic ambiguity resolution using Spanish RCs as experimental stimuli. Prosodic breaks were manipulated in terms of presence and position. Findings from the three experiments showed that participants used prosodic boundaries as cues in their interpretation of the stimuli, and that prosodic breaks were found to alternate participants' preferences for RC attachments.

2.2.2.4. Garden Path Sentences

Fodor (2002), Wagner and Watson (2010) treated a classic example of garden path sentences, as shown in (42), as one of the locally ambiguous constructions which cannot be resolved prosodically.

(42) The horse raced past the barn fell.

The preferred interpretation of a garden path sentence generally conforms with the Minimal Attachment and Late Closure Principles (Frazier, 1979) which suggest that comprehenders are likely to attach the incoming material to the syntactic constituent(s) that is/are being processed. This attachment consequently leads to participants' incorrect reading of the sentence during early processing. Reanalysis of the sentence is expected to occur when the comprehender notices difficulty in the processing of the sentence.

(43) When John was reading the book dropped onto the floor.

The Late Closure Principle predicts that processors will attach the NP *the book* to the most recent constituent *reading* rather than treating it as the phrase that starts the new constituent (van Gompel, Pickering, Pearson, & Jacob, 2006). In example (43), *the book* was initially analyzed as the direct object of the verb *reading*; in other words, *the book* was initially mistakenly to be part of the VP of which the head is *reading*. The presence of the verb *dropped* requires attachment of the verb to the preceding subject, making the initial reading become irrational. When processing difficulty arises, reanalysis of the sentence will take place.

Garden Path theory proposed that comprehenders do not wait until the end of the sentence to syntactically analyze the sentence, but rather adopt a serial, parallel approach to syntactic parsing (van Gompel, Pickering, Pearson, & Jacob, 2006). Upon encountering a garden path sentence, parsers are predicted to instantly analyze the current constituent while

simultaneously proceeding to the subsequent constituents of the sentence. Reanalysis of the sentence takes place immediately after the parser experiences difficulty in processing. This hypothesis claims that comprehenders adopt a processing mechanism which concurrently exploits a wide range of information to guide the interpretation of the sentence. Kjelgaard and Speer (1999); Marslen-Wilson, Tyler, Warren, Grenier, and Lee (1992); and Speer, Kjelgaard, and Dobroth (1996) further investigated the reciprocal influence between prosodic grouping and the Late Closure Principle and yielded similar findings supporting the use of prosodic information during early processing.

Though the above-mentioned studies generally claimed that garden path sentences such as (42) and (43) could not be prosodically resolved (e.g., Fodor, 2002; Wagner & Watson, 2011), contrasting viewpoints also emerged from other research. Grillo, Aguilar, Roberts, Santi, and Turco (2018) argued that garden path sentences can still be prosodically disambiguated when the sentence is embedded within a matrix clause which provides "a baseline pace" (p. 3) for the change in prosodic contour to take place. Grillo et al. (2018) designed 16 experimental stimuli for each of the two conditions: (a) main verb condition; and (b) reduced relative clause condition. The first condition differed from the second one in the insertion of the coordinator *and* which functioned as the disambiguating coda.

(44) (a) Jason claims that the student pushed into the row of traffic and got badly hurt.

(b) Jason claims that the student pushed into the row of traffic got badly hurt.

Participants of the experiments were instructed to read the stimulus silently, then verbally produce the sentence at normal speed. Findings from the experiment showed that the stimuli in the main verb condition were uttered at regular pace, whereas the sentences in the embedded RC condition were produced at faster reading speed. These results indicated the

difference in the prosodic contours between the two conditions, suggesting that garden path sentences could be prosodically disambiguated. It was further inferred from the results of the experiment that comprehenders process complex syntactic structures such as embedded reduced RC constructions faster than simpler syntactic constructions. These findings present the need for future research to shed light on the processing of garden path sentences embedded in a matrix clause.

As reviewed at the beginning of section 2.2, findings concerning the effects of intonational breaks on resolution of syntactic ambiguity were not consistent among past research. Such inconsistency was in part due to the use of experimental sentences that differ in syntactic structures. Moreover, different types of sentential ambiguity require different disambiguation strategies. For semantic ambiguity (e.g., *Flying planes can be dangerous.*), the prosodic contour of the sentence did not seem to provide any help in disambiguation. In addition, during real time processing, memory also contributes to the resolution of ambiguity, and under certain circumstances may function as covariates with prosody in syntactic disambiguation.

2.2.3. Prosody and Working Memory

Research has shown that working memory capacity plays a significant role in the parser's ability to retain information in memory to accurately retrieve the cues that guide interpretation (Gibson, Desmet, Watson, Grodner, & Ko, 2005; Grodner & Gibson, 2005; Swets, Desmes, Hambrick, & Ferreira, 2007). Constraints on working memory are reflected in participants' task performance in experiments that are specifically targeted at measurements of memory capacity and the effects of memory capacity on comprehension and production. Theories on sentence processing suggested that processing difficulties emerge when

comprehenders fail to sufficiently retain information about the target dependencies in memory (Abney & Johnson, 1991; Chomsky & Miller, 1963; Gibson, 1991, 1998; MacWhinney, 1987; Miller & Isard, 1964; Pickering & Barry, 1991; Stabler, 1994; Yngve, 1960).

Frazier's (1978) Minimal Attachment Principle was based on the effects of memory constraints, which associates comprehenders' minimal attachment preference for complex syntactic structures with limited working memory capacity. Research that featured garden path sentences associated cognitive burden with comprehenders' initial misinterpretation of the experimental sentences. The cognitive load is hypothesized to originate from the comprehenders' failure to simultaneously compute multiple syntactic structures during processing. Syntactic analysis during early processing requires the storage of unanalyzed syntactic constituents in memory, which consequently takes a toll on memory (Swets et al., 2007).

Gibson et al. (2005), and Grodner and Gibson (2005) stated that sentence processing is constrained by the limited amount of information comprehenders can store in memory. Lewis and colleagues (1996) developed a processing model, suggesting that sentence parsing involves the connection between heads and their corresponding dependencies in higher hierarchical positions. Such an approach to processing recognizes working memory as a decisive factor in parsing and comprehension. In regard to online sentence processing, the degree to which different comprehenders deploy different sources of information during comprehension, and the extent to which comprehenders are affected by different processing constraints were claimed to be in relation with the capacity of working memory (Swets et al., 2007).

For garden path sentences, there are contrasting accounts concerning the type of information that processors utilize at the early stage of processing. Frazier (1987) stated that a

comprehender only makes use of grammatical information upon his/her first encounter of the target discourse segment, other types of information will be taken into consideration by the comprehender during later stages of processing. Opponents to this account (e.g., MacDonald, Just, & Carpenter, 1992; Pearlmutter & MacDonald, 1995) argued that the amount and the types of information a comprehender utilizes during processing depends on his/her working memory capacity. The higher the memory capacity is, the larger the amount of information a comprehender can store and maintain in memory.

Swets et al. (2007) examined the effects of working memory on participants' decisions on attachment in syntactically ambiguous relative clauses. The experimental stimuli of the research were designed in English and Dutch, and distributed respectively to the two groups of participants (i.e., English and Dutch natives). Findings from the experiments revealed that when the stimuli (e.g., *The maid*$_{NP1}$ *of the princess*$_{NP2}$ *who scratched herself in public*$_{RC}$ *was terribly embarrassed*.) were presented as a complete sentence, participants having lower memory capacity were more likely to divide lengthy discourse chunks into smaller segments since their memory could not handle large chunks of text, which resulted in their HA preference of the relative clause. However, when the stimuli were presented as smaller segments with intended breaks inserted purposefully between NP_1, NP_2, RC, and the matrix VP, participants in both language groups biased towards HA preference regardless of the difference in their working memory capacity. Swets et al. (2007) interpreted such findings as the effects of implicit prosody aligning with the line breaks inserted in the stimuli on RC attachment decisions.

Traxler (2007, 2009) reported similar results from eye-tracking experiments, showing the influence of prosody on attachment preferences. It was observed that the line breaks placed between segments of the stimuli helped guide the participants' decisions on RC attachments.

Findings from these two studies, among some others, suggested that implicit prosody plays an important role in sentence processing, which will be further discussed in section 2.2.4.

The effects of prosody on language processing lead to the hypothesis that manipulating prosodic grouping in a sentence (or an utterance) will consequently tax working memory, especially in tasks that require complex syntactic parsing. As a result, the disruption or exclusion of prosody from a sentence/utterance, to a certain extent, will affect the retrieval of cues held in memory, placing a greater burden on comprehenders' cognitive processing. Pratt and Fernández (2016) introduced a processing model that integrated prosody and cue-based memory retrieval into the hypothesis of good-enough processing (Ferreira, 2002; Ferreira et al., 2003). The model was suggested to be applied to all language users, irrespective of their language profiles. With respect to the role of implicit prosody, findings from Pratt and Fernández's (2016) research confirmed that manipulations of implicit prosodic contours affected participants' performance in comprehension accuracy tasks.

2.2.4. Prosody in Reading

Empirical research has shown a positive correlation between prosody and reading skills (e.g., Kocaarslan, 2019) in both oral and silent reading. To gain a more profound understanding of the relationship between prosody and reading comprehension, it is important to look into oral reading fluency and the interaction between prosody and syntactic analysis during silent reading.

2.2.4.1. Prosody in Oral Reading

Findings from earlier research have pointed to an association between prosodic grouping and reading fluency. Dowhower (1991) stated that the nature of prosody is syntactically linked with the grammatical structure of a sentence. In order to achieve reading fluency, the reader is

required to have not only semantic and syntactic knowledge, but also accumulated experience in reading different genres of texts. The richer the reading experience is, the more likely and rapidly it is for the reader to improve reading fluency. A reader's reading fluency is evaluated based on accuracy, oral reading rates, and prosodic parsing strategies (Fuchs, Fuchs, & Maxwell, 1988; Nathen & Stanovich, 1991; as cited in Pratt, 2015, p. 53).

Reading fluency consists of the following components: accuracy in decoding of words, automaticity in word recognition, and appropriate use of lexical stress and prosodic parsing (Dowhower, 1991; Kuhn & Stahl, 2003; Samuels, Schermer, & Reinking, 1992). Achievement in reading fluency depends on how well a reader performs each of the aforementioned tasks which occur successively during reading comprehension. Word decoding is required at the initial stage of processing, followed by automatized recognition of words. Subsequent tasks such as projection of a prosodic contour onto the text in alignment with the text's syntactic structure require higher level of reading skills. Moreover, performance of subsequent tasks also depends on the reader's skillfulness in the execution of the preceding tasks.

Concerning the relation between fluency and comprehension, the two principal theories: Automaticity theory (LaBerge & Samuels, 1974) and Verbal Efficiency theory (Dowhower, 1991; Kuhn & Stahl, 2003) respectively focus on cognitive-based processing (i.e., automaticity in reading) and the role of prosody in fluency development. Under the Automaticity theory, reading is regarded as a complicated process which requires the reader to perform multiple tasks simultaneously. In addition to decoding and recognizing words, the reader must be able to interpret and cohesively gather the intended meaning of each individual word given the context in which the words are used. Simultaneous performance across multiple tasks requires automaticity of low-level processes such as word decoding, word recognition or lexical access

so that the reader can spare processing capacity for other higher level processes such as semantic derivation, syntactic parsing, and comprehension. Failure to automatize low level processes might result in the cost of comprehension (LaBerge & Samuels, 1974; Stanovich, Cunningham & Feeman, 1984). Nevertheless, other researchers (e.g., Dowhower, 1987; Taguchi, Gorsuch, & Sasamoto, 2006) have expressed concerns about Automaticity theory, particularly the automatization of low-level processes such as word decoding and word recognition. Opponents to the Automaticity theory found that improvement in word decoding and word recognition does not necessarily correlate or contribute to the improvement of reading fluency and comprehension.

Though automaticity of word decoding and word recognition during the earlier stage of processing was supported by the Verbal Efficiency theory (e.g., Dowhower, 1991; Kuhn & Stahl, 2003), automatization in low-level processes alone is not enough for reading fluency to be achieved. Automatized high-level processes, as argued, plays a larger role in enhancing reading skills by improving both reading speed and comprehension. High-level processes include the ability to parse discourse into smaller yet meaningful segments (i.e., verbally projecting syntactic boundaries onto the text). Evidence from empirical research also showed that when readers are given relevant semantic and syntactic information, their comprehension of the text will be significantly improved (Miller & Schwanenflugel, 2006).

Variations in reading fluency and comprehension accuracy were observed among fluent and non-fluent readers. Fluent readers are more skillful in chunking texts into appropriate meaningful segments, thus finding reading relatively effortless. Non-fluent readers tend to read texts on a word-by-word basis. In cases when non-fluent readers manage to group individual words into larger chunks, the grouping is more likely to run counter to the normal syntactic

and/or prosodic parsing of the text (Kuhn & Stahl, 2003); henceforth, does not contribute much to processing or comprehension.

As prosody consists of the following components: pausal intrusions, length of phrases, appropriateness of phrases, phrase-final lengthening, terminal intonation contours, and stress (Dowhower, 1991), projecting prosodic contours onto a text is viewed as acoustically parsing the text into smaller chunks, each of which has a syntactic role corresponding to that of the word/phrase in the text. Absence of prosodic information in a written text might pose difficulties to readers in processing and comprehension (Schreiber, 1987).

Understanding the importance of text segmentation in sentence processing, a number of studies have been conducted to shed light on the effects of implicit prosody on sentence processing (e.g., Pratt & Fernández, 2016; Yamashita & Ichikawa, 2010). Manipulations of implicit prosody in the form of text presentation formats are expected to affect processing.

Presentation formats that align with the syntactic parse of a sentence are predicted to reduce cognitive load; thus, facilitate processing. On the contrary, if the text is prosodically segmented following a pattern which deviates from the syntactic boundaries of the text, comprehenders may be expected to experience disruptions in processing (Yamashita & Ichikawa, 2010) as shown in example (69) in section 2.3.5.3. Gerber-Moron, Szarkowska, & Woll (2018) also examined the effects of text segmentation on participants' processing of video subtitles. The subtitles were intentionally manipulated to either follow or violate the syntactic parsing of the discourse. Though some participants in the study were audio impaired, findings from the study were consistent with findings from preceding research, indicating that text segmentation formats might either reduce or increase cognitive load, especially across tasks that require detailed syntactic analysis of the sentence.

Earlier research investigating the effects of text segmentation on reading comprehension focused on the development of reading fluency by both children and adults (e.g., LeVasseur, Macaruso, Palumbo, & Shankweiler, 2006; O'Shea & Sindelar, 1983; Rasinski, 1990). Findings from those studies showed that segmentation of a text into smaller units contributes to improvement in reading performance in both children and adults with poor to average reading skills (e.g., Cromer, 1970). In contrast, higher proficient readers were not significantly affected by manipulations of text segmentation; instead, they managed to use their syntactic knowledge to identify typical prosodic breaks of the text (Cromer, 1970; Rasinski, 1989). Results from those studies indicated that text segmentation has different effects on comprehenders during reading, depending on the reading proficiency of the comprehender.

Pratt and Fernández (2016) found that L1 and L2 speakers of English are differentially affected by manipulations of text segmentation. The experimental sentences in Pratt and Fernández's study were presented in one of the following formats: word-by-word; phrasal segment, or whole sentence, as shown in (45).

(45) The coach | who loved the players on the soccer team | encourage(s) everyone before each big game.

While the phrase-by-phrase presentation format was found to be facilitative to L1 participants across experimental tasks, the format did not benefit the other group. With respect to L2 reading fluency, an in-depth review of text segmentation in L2 reading will be provided in section 2.3.5.3. It is expected that the manipulations of prosody in the form of text segmentation formats will differentially affect L1 and L2 participants in the processing and resolution of reflexive-antecedent dependencies.

2.2.4.2. Prosody in Silent Reading

Functions and effects of prosody in speech have been well recognized in empirical research. Prosodic contours are predicted to contribute to resolution of syntactic ambiguity (e.g., Kjelgaard & Speer, 1999; Snedeker & Trueswell, 2003), relative clause attachments (e.g., Pratt & Fernández, 2016), polysemantic words (e.g., Schafer, Speer, Warren, & White, 2000), and sarcastic/irony expressions (e.g., Nakassis & Snedeker, 2002). Absence of prosodic contour in the input might have a detrimental effect on comprehension. As reviewed in 2.2.4.1, prosody has been argued to affect retrieval of information in oral reading (Koriat, Greenberg, & Kreiner, 2002), such that appropriate segmentation of syllables within a word, and segmentation of phrasal units within a sentence contribute to accurate interpretation of the sentence. In silent reading, integration of prosody into sentences is associated with improvement in reading comprehension (Miller & Schwanenflugel, 2008). Disruption of prosody or monotonous rendering of texts increases cognitive load, which might result in sentence misinterpretation.

A significant amount of earlier work focused on the role of prosody in children's reading skills and comprehension (e.g., Miller & Schwanenflugel, 2008; Schwanenflugel, Westmoreland, & Benjamin, 2015). Young children are claimed to initially make use of prosody in oral reading, then accordingly apply prosody in silent reading. Sensitivity to linguistic focus in oral reading was also observed among young readers. Those children were found to demonstrate linguistic sensitivity to the acoustic contour of a sentence, which was demonstrated in their change of pitch and acoustic intensity to match the syntactic parsing of the sentence.

Over the years, a number of hypotheses have been formulated with respect to the production of prosody in reading. Koriat et al. (2002) introduced The Structural Precedence hypothesis and claimed that readers are first guided by lexical and morphosyntactic cues during

early processing. The analysis of a sentence's syntactic structure, as proposed by Koriat et al. (2002), is performed by readers without advance preparation, and was argued to be "relatively independent from semantic information" (p. 272).

Alternative accounts, on the other hand, emphasized the effects of prosody during both early and late syntactic analysis (re-analysis) of a sentence. Bader (1998) and Fodor (2002) suggested that projection of implicit prosody onto written texts facilitates readers in syntactic parsing including parsing decisions occurring during the stage of syntactic re-analysis. Evidence supporting Bader's (1998) and Fodor's (2002) findings predominantly came from studies on ambiguous sentences or garden path sentences. Prosodic parsing helps resolve syntactic ambiguity such as (46) as the prosodic breaks correspond to the syntactic structure of the sentence. PP attachment preferences (HA or LA) in (46) are expected to manifest in the readers' decisions on prosodic phrasing.

(46) The man greeted the woman with a flower.

The reanalysis of a garden path sentence requires readers to take into consideration both the prosodic structure and the syntactic structure of the sentence to guide themselves towards the final interpretation of the sentence (Bader, 1998). Preferences for PP attachments, NP attachments, and RC attachments are also determined by readers' working memory capacity. Findings from Swets et al.'s (2007) research showed that high working memory capacity is associated with preferences towards LA, whereas HA is favored by participants with low working memory capacity. Observations from the processing of syntactic ambiguity and garden path sentences paved the way for the establishment of the Implicit Prosody hypothesis (Fodor, 1998, 2002) which suggests that the projection of prosody onto texts in silent reading influences readers' interpretations of syntactically ambiguous sentences.

Implicit prosody is regarded as an expressive inner voice in silent reading that readers internalize in their head as they silently read along a passage. The association between prosodic reading and fluency has also been well received by empirical research. Nevertheless, the application of implicit prosody in silent reading and the transition of prosody from spoken speech to written texts still call for further research. Syntactic parsing has been vindicated to frequently align with prosodic rendering, and thus, to some extent, forecasting the prosodic grouping of a written text. However, the role of syntax and punctuation in readers' decisions on the prosodic segmentation of a given text has still been under-examined. Punctuation in both oral and silent reading was proposed to influence readers in comprehension. Steinhauer (2003) found that punctuation corresponds with an inner voice via the observation of similar online brain responses between commas and speech boundaries. Conclusions from Steinhauer's (2003) research indicated that the commas presented in a text function as facilitators guiding readers in phonological parsing. Similar results regarding the correlation between prosodic phrasing and punctuation were obtained from Kalbertodt, Primus, and Schumacher's (2015) study. Findings from their experiments showed that in addition to semantic and syntactic information, punctuation also plays an important role in prosodic phrasing. Kalbertodt et al. (2015) suggested that punctuation, prosody, syntax and discourse semantics should be treated as "independent but interacting domains with correspondence constraints between them" (p. 11).

Application of implicit prosody in silent reading was observed to vary across readers with different levels of reading skills. Processing of phonological units was found to occur at the early stage of lexical access (Ashby, Treiman, Kessler, & Rayner, 2006). Research on brain activity revealed that upon the initial visual encounter with a word, readers tend to immediately associate the phonological features of the word with the orthographic symbols by which the

word is represented. That indicated the guiding function of phonological coding in lexical access. In order to successfully establish a connection between the reading of a sentence with the intention of the writer, skilled readers were found to rely on word stress, prosodic parsing, punctuation, and other acoustic features of the sentence. Evidence from earlier research suggested that skilled readers proactively make use of prosodic patterns to distinguish the contrastive information between the utterance they produce and the input they have received earlier (Cooper, Eady, & Mueller, 1985). Hellbernd and Sammler (2015) found that the intentions of the speaker are demonstrated via the prosodic contour of the sentence. In oral reading and communicative discourse, the correct interpretation of the prosodic signal produced by the speaker partly contributes to the success of interpersonal communication (Hellbernd & Sammler, 2015). Bridging that to silent reading, skilled readers who can extract the prosodic information from a sentence are expected to gain a thorough understanding of the writer's intention embedded in the text. Groen, Veenendaal, and Verhoeven (2019) investigated the performance of participants differing in age and reading skills on speech prosody. Results from their research indicated that poor readers showed deficiency in both perception and production of prosody. Groen et al. (2019) suggested that prosodic information plays a crucial role in the formation of a relevant structural representation of written texts.

Regarding the association between implicit prosody and the processing and resolution of syntactic dependencies, much research has been dedicated to investigating subject-verb agreement. Kreiner (2005) concluded from the findings of their eye-tracking experiments that participants' sensitivity to violations of subject-verb agreement depends on the prosodic phrasing of the experimental sentences. While natural prosody facilitates detection of agreement violations, disruptions in prosodic contours, on the other hand, hinder processing and negatively

affect participants' ability to track down violations of subject-verb agreement. Emerging evidence from psycholinguistic research also suggested that rapid serial visual presentation of text (i.e., RSVP) such as word-by-word presentation tends to disrupt general reading comprehension (Bernard, Chaparro, & Russell, 2001; Kang & Muter, 1989; Pratt & Fernández, 2016).

To the date of the present study, the role of prosody in the resolution of anaphor-antecedent dependencies has not received much attention from research. Wolters and Byron (2000) examined the effects of prosodic marking on the resolution of pronouns. Results from their research showed that prosodic information is not a strong indicator for the prediction and interpretation of anaphor-antecedent. Findings from Klassen and Wagner's (2017) study suggested that prominence shifts indicate anaphoric relationships to contextual antecedents similar to pronouns. Prosodic prominence was argued to correspond with speakers' intentions which were manifested through the use of different anaphors. Since the effects of prosody on real time resolution of reflexive anaphors have not been extensively studied by empirical research, this study would like to fill in such a gap in the literature.

2.3. Second Language (L2) Processing

A large body of L2 processing research has focused on the mental processes, mechanisms, and architectures involved in second language acquisition. A nonnative speaker, as a result, is considered a language processor who relies on his/her processing capacity to acquire and use a second language. Initial studies on L2 speakers' mental processes are suggested to evolve from Cattell's (1887) research which investigated the duration bilinguals spent on picture-naming tasks using both their first and second language. Results from Cattell's study showed that it took participants more time to name the pictures in the second language as

compared to their first language. The amount of time participants spent on L2-L1 translation was longer than the time needed for the L1-L2 translation. Similar findings generated from follow-up research in the 1980s provided further evidence of the discrepancies concerning cognitive load in language processing among L1 and L2 speakers (e.g., Flege & Hillenbrand, 1984; Killborn & Cooreman, 1987; Koda, 1989).

Different theoretical models and frameworks of L2 processing have been developed over the past decades in addition to the growing number of experimental studies to offer further insights into the nature and the issues of L2 processing. Earlier research on L2 processing geared towards: (i) L2 lexical processing (e.g., Koda, 1989), (ii) L2 phonological processing (e.g., Flege & Hillenbrand, 1984), and (iii) L2 sentence processing (Killborn & Cooreman, 1987). Those three aforementioned research areas have still remained the focus of attention in the current research of L2 processing. Though the three research areas deal with distinctive linguistic aspects, their research aims center around the followings: association between L1 linguistic knowledge and L2 processing; L1 transfer in L2 processing; L2 learning backgrounds including L2 proficiency, age of onset, years of residence in a country where an L2 is officially used as the first language, etc.

Since the primary focus of L2 processing is on mental processes, the dominant research areas in L2 processing are cognitive psychology, psycholinguistics and neurolinguistics. The majority of the theoretical frameworks relevant to L2 processing and research methods are rooted from psycholinguistic research on L1 processing. In other words, psycholinguistic studies on L1 acquisition provided the starting point for research on L2 processing. Nevertheless, L2 processing, by nature, still has unique and distinctive features, which helps contribute to the recognition of L2 processing as an extended but independent area of cognitive research. L2

processing has been treated as a constitutive part of second language acquisition (i.e., SLA) research since cognitive processing is part of the acquisition process (Clahsen & Felser, 2006).

2.3.1. L2 Syntactic and Morphosyntactic Processing

One of the three main components in L2 processing is sentence processing which involves comprehenders' parsing strategies that guide the interpretation of a sentence. Given the focus on online sentence processing, this section centers on the review of L2 processing, particularly L2 processing of (i) syntactic ambiguity, (ii) morphosyntactic features, and (iii) anaphor-antecedent dependencies.

2.3.1.1. L2 Processing of Syntactic Ambiguities

There has been an increased interest in L2 parsing strategies in the processing and resolution of syntactic ambiguities that involve attachment of relative clauses (RC) and prepositional phrases (PP). Speakers of different languages have been observed to have cross-linguistic variations in parsing strategies during the processing and resolution of syntactically ambiguous sentences that contain RC and PP attachments.

2.3.1.1.1. RC Attachments

As for resolution of RC attachments, English language speakers (e.g., Carreiras & Clifton, 1999) and speakers of languages such as Norwegian, Romanian, and Swedish (e.g., Ehrlich, Fernández, Fodor, Stenshoel & Vinereanu, 1999) were observed to have preferences towards low attachment. Whereas, high attachment preferences were found among participants of languages such as Japanese (e.g., Kamide & Mitchell, 1997), German (e.g., Hemforth, Konieczny, & Scheepers, 2000a), Dutch (e.g., Brysbaert & Mitchell, 1996), Spanish (e.g., Carreiras & Clifton, 1999), and Russian (e.g., Sekerina, 1997).

In addition to earlier processing theories and processing models of syntactic ambiguity such as the Garden Path model (Frazier & Fodor, 1978), and the Minimal Attachment and Late Closure principles (Frazier, 1978), other models have been developed over the past decades to account for the cross-linguistic differences in syntactic ambiguity resolution, among which is the Construal hypothesis (Carreiras & Clifton, 1993; Frazier & Clifton, 1996).

Frazier and Clifton (1996) developed the Construal hypothesis as a modified version of the Garden Path theory. The Construal hypothesis suggests that comprehenders exhibit their HA or LA preferences of the RC based on the two parsing decisions: *primary* and *non-primary* phrases. A primary phrase is a constituent that is predicted by the recent computation of the syntactic unit in which the primary phrase is embedded. In other words, if a complementizer phrase (CP) has already been computed, a verb phrase (VP) embedded in such a CP is regarded as a primary phrase. In contrast, a non-primary phrase is any constituent which is not syntactically mandatory, such as an adverbial phrase. The Construal hypothesis states that the processing of primary phrases is governed by the Garden Path theory while the processing of non-primary phrases is not. As non-primary phrases such as RCs are not processed using the Late Closure principle, the Construal hypothesis suggests that the RC is associated with the current thematic processing domain which can either be the high or the low NP of the ambiguous sentence as illustrated in example (41).

Comprehenders' decisions on high or low attachment of the RC depends on which NPs that the comprehender considers "referential". The Referentiality principle (Gilboy, Sopena, Clifton, & Frazier, 1995) proposes that the head of the current thematic processing domain (i.e., maximal projections of the last thematic role assigner) that introduces new discourse entities into the context is referential. Following the Referentiality principle, the head *the servant* of the

NP *the servant of the actress* in example (41) is referential, and thus is more likely to be perceived by comprehenders as being modified by the RC *who was on the balcony*. Nonetheless, as English language speakers have been found to have preferences for RC low attachment, explanations to this observation have been hypothesized to trace back to the origin of the English language which was, at the time, influenced by both the Saxon and Norman languages. Constructions of genitives in Saxon and Norman show different preferences for attachment. In cases of syntactic ambiguity, languages that have both of the two types of genitive constructions (i.e., Saxon and Norman) such as the English language tend to gravitate towards the Norman genitive constructions which favor low attachment to avoid syntactic ambiguity (Dinçtopal-Deniz, 2010).

The other RC attachment preference hypothesis that is similar to the Construal hypothesis is the Recency and Predicate Proximity proposed by Gibson et al. (1996) and updated by Pearlmutter and Gibson (2001). Gibson et al. (1996), and Pearlmutter and Gibson (2001) suggested that the decision on the attachment of the RC in a syntactically ambiguous sentence is determined by the two competitive factors: the Recency Preference which favors low attachment, and the Predicate Proximity which gives more preferences to high attachment. The Recency Preference is suggested to be very close to the Late Closure Principle (Frazier, 1978) as the two principles both predict a preference towards LA due to constraints on comprehenders' memory capacity. Gibson et al.'s (1996) Recency Preference predicts that languages with a strict word order such as English are more likely to show preferences towards LA in accordance with the relatively close distance between the verb and its corresponding arguments in English. On the contrary, languages such as Spanish, Russian, German, and

French that syntactically allow lengthier distance between the verb and the arguments are observed to display a HA preference.

Given the similarities between the Construal hypothesis (Frazier & Clifton, 1978) and the Recency/Predicate Proximity (Gibson et al., 1996, 2001), Dinçtopal-Deniz (2010) investigated the processing of RC ambiguity by native English speakers and Turkish speakers of English and discussed the results in relation to the two above-mentioned hypotheses and the Shallow Structure hypothesis (Clashsen & Felser, 2006) (The Shallow Structure hypothesis will be reviewed in section 2.3.3.1). Findings from Dinçtopal-Deniz's (2010) study indicated that L1 participants showed stronger preferences towards LA regardless of the animacy feature of the antecedents, while L2 participants showed variations in their preferences for RC attachments across online and offline tasks, and across animate and inanimate antecedents.

Inconsistent findings regarding L1 and L2 preferences for RC attachments were also observed in other research (e.g., Dussias, 2003; Papadopoulou & Clahsen, 2003). Participants in Dussias's (2003) research showed a bias towards LA irrespective of their L1 (either English or Spanish).

(47) (a) A man called the student$_{(+masc)}$ of the teacher$_{(+fem)}$ who was disappointed$_{(masc)}$ by the new educational system.

(b) A man called the student$_{(+masc)}$ with the teacher$_{(+fem)}$ who was disappointed$_{(masc)}$ by the new educational system.

Papadopoulou and Clahsen (2003) found variations in participants' preferences for RC attachments in different types of RC syntactic ambiguity. While only the L1 (Greek natives) favored a high attachment of a RC containing the genitive *of*, both the L1 and the L2

participants whose first languages included German, Russian, and Spanish showed LA preferences for attachment of the RC containing the preposition *with* as shown in (47).

2.3.1.1.2. PP Attachments

Research on L2 processing of PP attachment ambiguities has extended to different languages such as English (e.g., Felser, Roberts, Marinis, & Gross, 2003), French (e.g., Frenck-Mestre & Pynte, 1997), Greek (e.g., Papadopoulou & Tsimpli, 2005b), etc. Frenck-Mestre and Pynte (1997) examined the parsing strategies applied by L1 and L2 speakers of French in the resolution of syntactically ambiguous sentences containing PP such as (48).

(48) Brutus hit the gladiator ***with the shield*** with his bare hands.

The domain of syntactic ambiguity of the experimental sentence closes at the end of the PP *with the shield*, which allows two possible readings of the sentence depending on the attachments of the PP to either the VP *hit* or the DP *the gladiator*. Following the Minimal Attachment Principle of the Garden Path hypothesis (which gives more preferential weight to an NP-attachment of the PP), the PP *with the shield* is preferred to be attached to the DP *the gladiator*. In contrast, if parsing decisions are based on the lexical information of the verb *hit*, the attachment of the PP *with the shield* should be biased towards the verb. In order to get insight into whether parsing decisions are guided predominantly by lexical information or syntactic information, Frenck-Mestre and Pynte (1997) used experimental sentences such as (49) in the second experiment of the study.

(49) When the musician played (came) the white grand piano was in the centre of the stage.

Quand le musicien jouait (venait) le beau piano blanc eâ tait au centre de l'estrade.

Findings from the second experiment revealed that lexical information has certain effects on participants' syntactic parsing decisions. Though attachment of the PP to the DP was dominantly favored by L2 speakers, this parsing strategy was affected, but not completely governed, by the lexical information extracted from the verb. French-Mestre and Pynte (1997) also found effects of L1 influence on parsing decisions in L2 sentences.

Contrary to French-Mestre and Pynte's (1997) findings, Fernández's (2000) study which used self-paced reading experiments to investigate parsing preferences of Spanish and English bilinguals found no L1 influence on sentence processing strategies. Mixed results from different studies (e.g., Dussias, 2001; Papadopoulou & Clahsen, 2003) regarding L2 PP attachment preferences as a result of L1 transfer laid the foundations for subsequent research. Felser et al. (2003) examined resolution strategies employed by L2 speakers of English in the resolution of syntactically ambiguous sentences such as (50a) and (50b).

(50) (a) The young girl favored the driver of the player who was talking to an old woman.

(b) The young girl favored the driver with the player who was talking to an old woman.

L2 participants in the study were grouped into two groups based on their L1: Greek and German. Results from the four experiments indicated that L2 speakers differ from native English speakers in the processing of syntactically ambiguous sentences. Participants' preferences towards ambiguity resolution were affected by the lexical meaning of the prepositions *of* and *with*; nevertheless, L2 speakers' processing strategies were observed not to resemble the strategies applied by monolinguals. Felser et al. (2003) also found that children,

monolinguals and bilinguals are differentially influenced by lexical-semantic and syntactic information in online sentence processing and syntactic ambiguity resolution.

2.3.1.2. L2 Processing of Morphosyntactic Features

Research on L2 sentence processing, in general, has supported the association between comprehenders' grammatical knowledge and their decisions on syntactic parsing (e.g., Clahsen & Felser, 2006a). When a comprehender first arrives at the initial word of a sentence, he/she will immediately deploy his/her knowledge of grammar to analyze all the relevant grammatical constraints governing the lexical and syntactic use of the word so as to grammatically establish the connection between the words in the sentence. The detection of a grammatical violation will trigger the comprehender's reanalysis of the sentence. Methodological advances in psycholinguistic research have made it possible to detect sentence reanalysis, reflected in comprehenders' increased fixations and reading times, backward eye movements, or peaks in electrical brain activity (the P600) (Keating, 2009). Nevertheless, empirical research on L2 sentence processing has shown that sentence reanalysis induced by the detection of broken agreement in gender and/or number varies among L1 and L2 groups (e.g., Barber & Carreiras, 2005; van Berkum, Hagoort & Brown, 1999). Differences between L1 and L2 languages are typically attributed to the following: (i) the influence from L2ers' native languages which do not require gender and/or number agreement, leading to L2ers' insensitivity to gender/number agreement violations in the target language (Franceschina, 2001, 2005; Hawkins & Franceschina, 2004; McCarthy, 2008); or (ii) cognitive load from high task demands which has been suggested (particularly by Clahsen and Felser's (2016) Shallow Structure hypothesis) to hinder comprehenders' ability to make use of necessary resources to guide grammaticality judgments.

Keating (2010), and Sagarra and Herschensohn (2010) investigated the sensitivity L2 speakers of Spanish have towards violations of gender agreement. Results from Keating's (2010) study revealed that individual working memory capacity and the linear distance between the controller noun and the gender-inflected adjective differentially affect participants' sensitivity to gender agreement errors. These findings provided further evidence supporting Clahsen and Felser's (2016) Shallow Structure hypothesis. Sagarra and Herschensohn (2010) examined how adult L2 learners of Spanish with different Spanish proficiency levels responded to gender agreement violations, taking into account the animacy feature of the controller nouns. L2 learners with higher proficiency levels in Sagarra and Herschensohn's study were observed to be more likely to exhibit native-like sensitivity towards gender anomalies as compared to less proficient L2 learners. Findings from their research further showed that the animacy/inanimacy feature of the controller nouns also determines the variations in cognitive load placed upon L1 and L2 speakers of Spanish during online sentence processing.

In light of the effects of L2 proficiency levels on the processing of gender and number agreement, Alemán Bañon, Fiorentino, and Gabriele (2018) used ERPs to examine the L2 morphosyntactic development of English L1-Spanish L2 speakers with different proficiency levels of Spanish. There were 24 L1 Spanish speakers and 81 English L1-Spanish L2 speakers participating in the study. Alemán Bañon et al.'s (2018) research specifically looked at the following factors: (i) effects of proficiency on morphosyntactic processing in late bilinguals; (ii) proficiency measures that could be used to reliably predict L2 participants' sensitivity to morphosyntactic agreement; (iii) effects of L1-L2 similarities on morphosyntactic processing; and (iv) effects of hierarchical syntactic distance on L2 speakers' sensitivity towards morphosyntactic dependencies. Findings from their ERP experiments showed significant

differences in participants' ability to detect gender and number errors. The differences were most robust when comparing participants at the extremes of the proficiency range. Low proficiency participants were found to be sensitive to number agreement, reflected in the elicitation of a P600; however, the participants did not show any sensitivity to gender agreement errors. Alemán Bañon et al. (2018) also examined the influence of participants' native language on the processing of gender and number agreement in the target language. Though participants with lower Spanish proficiency were sensitive in detecting grammatical violations, their sensitivity was only restricted to violations of features that are also found in their native language. Sensitivity to other linguistic features such as structural hierarchical distance requires not only high levels of proficiency, but also gradual development of proficiency.

Subject-verb agreement has emerged as a focal area in behavioral and electrophysiological research (e.g., Osterhout & Mobley, 1995; Roehm, Bornkessel, Haider, & Schlesewsky, 2005; Wassenaar, Brown, & Hagoort, 2004) in which different ERP responses elicited during processing are suggested to indicate detection of subject-verb agreement violations. Given the influence of L1 on the processing of L2, Chen, Shu, Liu, and Zhao (2007) investigated Chinese L1-English L2 speakers' ERP responses to experimental constructions that either followed or violated subject-verb agreement in English. Seventeen native English speakers and 18 Chinese natives who had learned English for an average of 10 years were respectively recruited to participate in the study as a control group and an experimental group. The study used experimental stimuli targeting subject-verb agreement as shown in (51).

(51) The report from the ***agency/agencies was/were*** very encouraging.

Results from Chen et al.'s (2007) research showed that L2 participants with high proficiency in English were able to detect morphosyntactic violations in subject-verb agreement,

which was demonstrated in their performance on the grammaticality judgment task. However, upon encountering subject-verb agreement errors, L2 participants' brain responses showed distinctive patterns as compared to the brain responses elicited by the L1 group. The explanation given to this dissociation in the ERP responses between the two groups of participants was that the two languages (i.e., Chinese and English) require different morphosyntactic features in subject-verb agreement. In other words, syntactic features in English subject-verb agreement are absent in Chinese. Chen et al. (2007) concluded that the findings from their research were consistent with earlier research (e.g., Bates, 1999; Nelson, 1999; Zhang & Wang, 2007), that the grammatical differences between languages influence language processing and are reflected in differences in neural response patterns, and that neural functions and structures of the human brain required in language processing are language-dependent.

Armstrong, Bulkes, and Tanner (2018) also examined Chinese L1-English L2 speakers' ERP responses in subject-verb constructions similar to those used in Chen et al.'s (2007) research. Instead of manipulating the morphological plural marker of the first NP as Chen et al.'s, Armstrong et al.'s (2018) study varied the determiner standing before the first NP (e.g., ***The/Most*** *cookies* ***taste/tastes*** *best when dipped in milk.*). The verb in the experimental materials either morphosyntactically agreed or disagreed with the NP functioning as the subject of the sentence. Findings from the study showed that the Chinese L1 group produced a consistent P600 response to subject-verb agreement violations as compared to the English L1 group, which contrasts with the findings from previous research including Chen et al.'s (2007). The two groups of participants were observed to differ in their brain reactions towards constructions varying in the inclusion and exclusion of a quantifier in the subject NP. These observations

indicated that Chinese L1 participants relied on both lexical and morphological information to guide their decisions on grammaticality judgments on subject-verb agreement in English.

Taking a different approach, Pratt and Fernández (2016) investigated L1 and L2 processing of English subject-verb agreement in RC constructions. The researchers proposed a processing model that integrated prosodic contours and a memory retrieval mechanism into the hypothesis of good-enough processing. Findings from the self-paced reading experiments of the study demonstrated that L1 and L2 participants differed in their performance across grammaticality rating and comprehension tasks. Pratt and Fernández (2016) concluded that such divergence between the two groups of participants might be attributed to variations in text presentation formats, and fluency development in English reading.

With respect to L2 processing in English by a diverse group of L2 participants having different language profiles, Ghilzai (2017) investigated L1 influence on English sentence processing by speakers of Japanese, Persian, and Urdu. Findings from the study showed that regardless of the difference in their first language backgrounds, L2 participants displayed a more consistent pattern of accuracy and reaction time towards case anomalies than towards subject-verb agreement violations. The results indicated the absence of L1 influence on L2 sensitivity to morphosyntactic agreement in English.

2.3.1.3. L2 Anaphora Resolution

In recent years there has been a growing interest in L2 processing and interpretation of anaphors (e.g., Clahsen & Felser, 2006; Felser, Sato, & Bertenshaw, 2009). Research has shown that both syntactic and non-syntactic constraints influence L1 and L2 anaphora resolution; however, speakers of different language groups were found to be differentially affected by these factors. With respect to structural constraints guiding anaphora resolution, earlier studies

suggested that the application of binding constraints such as binding principles A, B, and C (Chomsky, 1981) was observed during the early stages of processing among native speakers, whereas L2 speakers were more likely to violate binding principles during early processing regardless of their levels of proficiency in the second language (e.g., Felser et al., 2009; Nicol & Swinney, 1989; Sturt, 2003).

2.3.1.3.1. L2 Processing and Resolution of Pronouns

Research has shown that sentence reanalysis can occur in L2 processing (e.g., Hopp, 2015; Roberts & Felser, 2011). These studies on L2 sentence processing measure reading times as a dependent variable. In cases of syntactically ambiguous sentences such as garden path sentences, increased reading times reflect participants' disambiguation of the sentence midway through processing (e.g., Dussias & Cramer Scaltz, 2008; French-Mestre & Pynte, 1997). Results from studies on L2 processing of syntactically ambiguous sentences (e.g., *While the band played the song pleased all the customers.*) indicated that though L2 participants were able to provide correct responses to the comprehension questions, their attempts to reanalyze the sentence to completely abandon the initial reading of the sentence were not consistently observed across experiments (e.g., Hopp, 2015; Jacob & Felser, 2006; Roberts & Felser, 2011; Slattery, Sturt, Christianson, Yoshida, & Ferreira, 2013). It could be inferred from these findings that compared to native speakers, nonnative speakers were more attached to their initial interpretation of the syntactically ambiguous sentence, and thus were less likely to recover from misinterpretation.

Cunnings, Fotiadou, and Tsimpli (2016) investigated the processing of English overt subject pronouns by Greek L1-English L2 speakers. Cunnings et al. (2016) examined whether L2 participants would reanalyze a syntactically ambiguous sentence when the subsequent

information of the sentence suggested another interpretation which stands in contrast with the initial interpretation of the sentence. Results from the experiments revealed that both groups of participants made use of discourse information (i.e., gender features) in the processing of subject pronouns. Though L2 participants were nativelike in the interpretation and processing of ambiguous English pronouns, they were more likely to linger on the first interpretation of the sentence than the L1 participants.

Patterson, Trompelt and Felser (2014) used eye-tracking to investigate the use of Binding Principle B in English L1 and English L2 (German L1) speakers. The aim of Patterson et al.'s (2014) study was to examine whether the application of Binding Principle B might rule out the likelihood of the inaccessible antecedent being considered for retrieval, and whether there was any L1-L2 difference in the application of the Binding Principle. The experimental sentences used in the three experiments were respectively as follows.

(52) The boy remembered that Matthew had bought him a new computer game. // Harry heard William pull the curtain around him in the quiet hospital ward.

(53) *John* remembered that *Mark* had taught *him* a new song on the guitar. // *John* remembered that *Jane* had taught *him* a new song on the guitar. // *Jane* remembered that *John* had taught *him* a new song on the guitar. That really lifted everyone's spirits!

(54) *Barry* saw *Gavin* place a gun near *him* on the ground with great care. // *Barry* saw *Megan* place a gun near *him* on the ground with great care. // *Megan* saw *Barry* place a gun near *him* on the ground with great care. The robbery was definitely over now.

Results from Patterson et al.'s (2014) experiments showed that native English speakers took into consideration the syntactic constraints governed by Binding Principle B during processing and resolution of both local and non-local pronouns. As a result, no interference effect from the inaccessible antecedent was observed among L1 participants, which is consistent with findings from earlier research on the resolution of anaphoric reflexives (e.g., Dillon et al., 2013). In contrast, L2 speakers were found to direct their pronoun interpretation preferences towards the antecedent which was more salient in discourse regardless of its syntactic position in the sentence. These findings provided further evidence concerning L1-L2 differences in the deployment of syntactic cues during processing and resolution of English pronouns.

Taking a similar approach, Slabakova, White, and Guzzo (2017) investigated the processing of reduced and full English pronouns by native adult speakers of French and Spanish. Both groups of participants were expected to encounter processing difficulties resulting from increased cognitive load, as suggested by Reinhart (2006). Slabakova et al. (2017) used a truth-value judgment task with experimental sentences such as (55).

(55) *Referential condition with the expected answer 'False'*:

Tom, Helen, and Harry were going to a soccer party. Prizes were going to be given out for the best spray-painted logo. They all sprayed the logo of their favorite soccer teams on their arms. Tom badly wanted to win the competition, so he asked his friends to help him make his logo even better. Helen refused to help because she wanted to win as well. Harry wanted to help Tom, but he had no spray-paint left.

Harry sprayed 'm. (Reduced pronoun experiment) T F

Harry sprayed him. (Full pronoun experiment) T F

Results from Slabakova et al.'s (2017) study showed that participants' performance on experimental tasks was affected by the type of pronouns and the antecedents used in the experimental materials. Participants' accuracy in the grammaticality judgment task decreased when full pronouns were accompanied by referential antecedents as compared to when reduced pronouns were used with quantified antecedents. The findings are in line with Reinhart's (2001) proposal concerning comprehenders' increased computational burden during pronoun processing and resolution.

2.3.1.3.2. L2 Processing and Resolution of Anaphoric Reflexives

Research on L2 processing of anaphors has mainly focused on the processing of anaphoric reflexives. Felser et al. (2009) investigated English L2-Japanese L1 speakers' sensitivity to binding principles during processing of English anaphoric reflexives. The first experiment was a timed/untimed grammaticality judgment task. Processing speed and interference effects were treated as outcome variables in the study. The experimental sentences in the first experiment contain violations of anaphor-antecedent agreement as shown in (56).

(56) (a) (Locality) **Mary* believed that the *dancers* had hurt *herself.*

(b) (C-command) *The *dancers* believed that *Mary's brother* had hurt *herself.*

Results of the first experiment indicated that L2 participants spent more time on the judgment task when there were locality violations. In other words, interference effects were more likely to occur when the inaccessible antecedent c-commanded the reflexive (as in 56a) as compared to when it did not (56b).

The second experiment was a speeded grammaticality judgment task. Experimental stimuli contained an introductory sentence which introduced the two referents (*John/Jane* and

Richard), and an experimental sentence in which the anaphoric reflexive was presented, as in (57).

(57) *John/Jane* and *Richard* were very worried in the kitchen of the expensive restaurant.

a. *John/Jane* noticed that *Richard* had cut *himself* with a very sharp knife.

b. It was clear to *John/Jane* that *Richard* had cut *himself* with a very sharp knife.

Kitchens can be dangerous places.

L2 participants' increased first-pass reading time indicated the interference effects from the distractor that matches the accessible antecedent in both gender and syntactic features. The finding suggested that the inclusion of an inaccessible antecedent that has a similar syntactic role as that of the accessible antecedent increases L2 participants' cognitive load. Felser et al. (2009) concluded that irrespective of native-like English proficiency, Japanese learners of English were negatively affected by the inaccessible antecedent which matched the target antecedent in discourse feature (interference effect). These findings are in contrast with findings from previous research on monolingual sentence processing (e.g., Sturt, 2003) which found no interference effects from inaccessible antecedents. Felser et al. (2009) suggested that syntactic constraints such as binding principles are not the only linguistic constraints that guide L2 processing and interpretation of anaphoric reflexives.

Based on Felser et al.'s (2009) research, Felser and Cunnings (2012) extended the investigation to the factors that predominantly affect L2 processing of reflexive anaphors. The study also looked into the time-course of the effects from both structural and non-structural constraints during processing. The L2 participants taking part in the experiments were native German speakers. The justification given by Felser and Cunnings for the recruitment of German

L1 speakers was that German is close to English in terms of syntactic features, such that none of the two languages favors long distance binding of reflexives. The inclusion of German native speakers in the study was assumed to eliminate L1 transfer which accounts for participants' confusion upon encountering items matching in syntactic features. The experimental materials used in the first experiment were adapted from Sturt's (2003) with gender agreement manipulated between the accessible antecedent, the inaccessible antecedent, and the anaphoric reflexive, as shown in (58).

(58) *James/Helen* has worked at the army hospital for years. *He/She* noticed that the soldier had wounded *himself/herself* while on duty in the Far East. Life must be difficult when you are in the army.

Results from the first experiments revealed that L1 and L2 speakers were differentially affected by structural and non-structural constraints in the processing of anaphoric reflexives. The native speakers were found to display similar patterns of processing as the monolingual English speakers in Sturt's (2003) research, such that Binding Principle A was applied during the early stage of processing. On the contrary, the L2 group, regardless of being nativelike in offline tasks, was observed to violate Binding Principle A during early processing, which was clearly demonstrated in their first-pass reading time of the reflexive.

(59) *James/Helen* has worked at the army hospital for years. The soldier that *he/she* treated on the ward wounded *himself/herself* while on duty in the Far East. Life must be difficult when you are in the army.

For the second experiment, Felser and Cunnings (2012) used experimental stimuli similar to those in Experiment 1, except that the inaccessible antecedent did not c-command the anaphoric reflexive, as presented in (59). The purpose of this experiment was to determine

whether binding constraints affect L2 participants in early processing of anaphoric reflexives. Results from the second experiment showed that L2 speakers directed close attention to the inaccessible antecedent during the initial stage of processing as a result of distractor prominence. L1 speakers were also observed to be affected by the interference effect from the inaccessible antecedent, which was demonstrated in the monitoring of their eye movement in the initial reading of the reflexive. However, this interference effect did not last long and did not extend to the post-critical region.

Findings from Felser and Cunnings's (2012) study demonstrated L1-L2 differences in the application of Binding Principle A in the processing of anaphoric reflexives: while L1 participants applied Binding Principle A immediately upon their initial encounter with the reflexive, L2 speakers – regardless of L1 influence – violated the principle and biased towards the prominent discourse feature of the inaccessible antecedent. Felser and Cunnings's findings are consistent with Clahsen and Felser's (2006) Shallow Structure hypothesis according to which syntactic processing is more compromised compared to general comprehension in L2 speakers.

To gain further insights into English L2 speakers' application of syntactic and non-syntactic information in the processing of anaphoric reflexives, Liang, Wen, and Dong (2018) used event-related potential experiments to investigate Chinese-English bilinguals' sensitivity to gender agreement between an anaphoric reflexive and its antecedent in Chinese and English. The first ERP experiment in Liang et al.'s research was a grammaticality judgment task that used Chinese experimental sentences in which the gender feature of the antecedent and the anaphoric reflexive was manipulated across two levels: *congruent* and *incongruent*. Experiment 2 examined how gender constraints affected the resolution of English pronouns in Chinese-

English bilinguals. The experimental stimuli used in Experiment 2 were adapted from Osterhout et al.'s (1997) where gender congruence/incongruence varied between the reflexive pronoun and its antecedent. Examples of the experimental sentences in Experiment 1 and 2 in Liang et al.'s study are respectively shown in (60) and (61).

(60) *Miss Zhao* found *herself/himself* in the dilemma.

(61) *Miss Wright* lost *herself/himself* in thought.

Findings from the two ERP experiments showed that the processing of Chinese and English anaphoric reflexives was guided by gender agreement between the reflexive and its antecedent. The processing mechanisms of the reflexive pronouns in the two languages were not exactly identical, as shown by different patterns of the P600 observed upon violations of gender agreement in English and Chinese.

Though the past decades have witnessed a huge growth in research on L2 processing of anaphoric reflexives, there has been little discussion on the effects of implicit prosody in L2 resolution of reflexive-antecedent dependencies in the literature. The following section reviews L2 processing and prosody in more depth, explaining why there is still a need for further studies that incorporate implicit prosody in L2 processing of anaphors.

2.3.2. L2 Processing and Prosody

Empirical research has generally found an interplay between syntax and prosody in L1 sentence processing (e.g., Beach, 1991, 1996; Cutler et al., 1997, 2002; Fraizer et al., 2006; Gee & Grosjean, 1983; Grodner et al., 2005; Katz et al., 1996; Price et al., 1991; Scott, 1982; Watson & Gibson, 2004). There has been growing interest in the interaction between syntax and prosody during processing with respect to L2 production and comprehension in both oral and silent reading (Fernández, 2010).

Prior to in-depth discussion on the association between syntax and prosody in L2 processing, it is important to look into the interplay between syntax and prosody in light of prosodic transfers. With respect to the effects of prosodic transfer in the production of L2 prosodic patterns, Nguyen, Ingram, and Pensalfini (2008) investigated how native Australian English speakers and Vietnamese speakers of English used word-level and phrase-level prosody to differentiate the three patterns of English stress: broad-focus noun phrase, narrow-focus noun phrase, and compound. Results from Nguyen et al.'s study showed that participants' native languages differentially affected their use of English acoustic features. In addition to L1 influence, L2 learning experience was also observed to have certain effects on L2 acquisition of English prosody. Vietnamese learners of English with low English proficiency were less likely to identify the association between prosodic and syntactic information in an utterance as compared to the native speakers. Low proficient speakers were also found to be more susceptible to difficulties in producing prosodic pitch patterns that guide structural information in the production task.

Goss and Nakayama (2011) investigated the production of prosody in oral reading in Japanese by English L1-Japanese L2 speakers. The study aimed to shed light on the relationship between comprehension and oral prosodic production. Goss and Nakayama predicted that errors in participants' prosodic production would indicate their poor comprehension of syntactic structures. Nonetheless, findings from their experiment proved otherwise. Errors made by Japanese L2 learners regarding oral prosodic parsing were not necessarily a consequence of deficiency in comprehension. The L2 participants in the study had high accuracy rates in the comprehension task despite their inability to place accurate prosodic breaks in the experimental materials.

Tsui, Tong, and Fung (2016) found contrasting results. The researchers examined the role of prosody in language production by Cantonese-English bilinguals. Participants of the study were asked to verbally produce six types of constructions varying in syntactic complexity. Results from the experiment revealed that bilingual children produced similar pitch patterns but different prosodic breaks as compared to adult native English speakers. The association between prosodic production and reading comprehension was reflected in participants' performance across experimental tasks.

In regard to the role of prosody in L2 sentence processing in languages which are not very linguistically isolated from one another, Fernández (2005) examined the use of prosodic information by early Spanish-English bilinguals in the production of English and Spanish syntactically ambiguous sentences. Association between reading comprehension and prosodic production was observed among the bilingual participants of the study. Dekydtspotter, Donaldson, Edmonds, Fultz, and Petrusch (2008); and Fultz (2007) extended the research to late bilinguals. Results from their studies suggested that even L2 participants with low level of L2 proficiency exhibited sensitivity to the incongruence between prosodic contours and syntactic parsing during processing of syntactically ambiguous sentences.

In a similar vein, Jackson and O'brien (2011) examined the use of prosodic cues by intermediate L2 learners of German in reading comprehension, and the types of prosodic information that L2 participants utilized in oral reading of syntactically ambiguous German sentences. The experiment of the study was a sentence production task in which a lead-in passage was followed by a syntactically ambiguous sentence with a PP either attached to the direct object or to the verb of the sentence, as shown in (62).

(62) (*NP-attachment*): Der geschäftsführer dankte der sekretärin mit zwei kindern. ("The manager thanked the secretary *with two children*.")

(*VP-attachment*): Der geschäftsführer dankte der sekretärin mit einem blumenstrauß. ("The manager thanked the secretary *with a bouquet of flowers*.")

Results from the experiment showed that L2 learners of German whose proficiency level was below advanced still made use of prosodic cues in the production task. L2 participants managed to syntactically parse ambiguous sentences to express their intended interpretation of the sentence. The findings provided further evidence to the proposal (e.g., Dekydtspotter et al., 2008; Fultz, 2007) that even low proficient L2 learners are sensitive to the association between the prosodic phrasing and the syntactic structure of a sentence, and that such interplay could function as a facilitator in sentence comprehension.

However, not much research has been conducted in regard to whether learners of different L1 backgrounds and L2 proficiency levels vary in the deployment of prosodic cues during processing and comprehension. In order to cast further light on this under-researched area, Nickels and Steinhauer (2018) used ERP experiments to examine the effects of L1 backgrounds (German and Chinese) and L2 proficiency on the processing of English garden path sentences (e.g., *When a bear is approaching the people (the dogs) come running*). As the first languages of the L2 participants of the study differ from English in the use of grammatical features (e.g., tenses and aspects), Nickels and Steinhauer expected to observe stronger preferences for late closure and early closure among L2 learners in the disambiguation of the garden path sentences. English proficiency was also treated as a factor affecting L2 participants' decisions and strategies in syntactic and prosodic parsing to resolve syntactic ambiguity since there was disagreement among researchers with reference to whether late bilinguals adopt the

processing architectures that resemble (e.g., Steinhauer, White, & Drury, 2009) or entirely differ from those applied by L1 speakers (e.g., Weber-Fox & Neville, 1996).

Findings from Nickels and Steinhauer's experiment are consistent with earlier research that indicated L1-L2 processing differences. L1 backgrounds were also observed to have an impact on L2 proficiency levels. As their mother tongue has distinctive linguistic features from the English language, the Chinese participants in the study generally had lower proficiency in English as compared to the German participants despite the fact that the Chinese group had L2 exposure in Canada while the German learned English as a foreign language in their home country. However, proficiency levels in L2 were found to override influences from L1 during processing of syntactic ambiguity. Provided that the participants were highly proficient in English, regardless of their L1 profiles, they could still elicit ERP patterns similar to L1 speakers in response to syntactic anomalies. Another note-worthy finding from Nickels and Steinhauer's study is that L1 reading proficiency also influenced native speakers in processing, which was demonstrated by the variations in L1ers' utility of prosodic cues in syntactic disambiguation.

To further examine the interaction between syntax and prosody in sentence processing, Pratt and Fernández (2016) investigated the effects of syntactic and prosodic information on L1 and L2 processing of subject-verb agreement in RC constructions. Pratt and Fernández (2016) manipulated the number feature between the subject and the verb of the experimental sentences. Implicit prosody was projected onto the experimental sentences via three different text presentation formats: word-by-word, phrasal segment, or whole sentence. Findings from the experiments showed that the two groups of participants were differentially affected by the three text presentation formats. L1 participants (i.e., English native speakers) benefited most from the

phrasal segment presentation. In contrast, L2 participants (i.e., Spanish L1-English L2) were not remarkably facilitated by the phrase-segment format and were even observed to perform poorly on the comprehension task in this paradigm. These findings pointed to L1-L2 differences in the effects of implicit prosody, supporting the crucial role of implicit prosodic parsing in the processing of subject-verb agreement.

2.3.3. Factors Affecting L2 Processing

As pointed out by previous research (e.g., Jackson & O'brien, 2011; Nickels & Steinhauer, 2018; Pratt & Fernández, 2016), L2 processing is affected by a number of different factors: L1 influence, L2 proficiency, working memory and task complexity. This section reviews findings from earlier research that touched upon the aforementioned factors.

2.3.3.1. L1 Influence

First language backgrounds might have different effects on L2 processing (e.g., Foote, 2011; Sagarra & Herschensohn, 2010; Sun, 2006). L2 learners are predicted to experience L1 interference during L2 processing, especially when the first language and the second/foreign language are linguistically distinct from one another, or at least different in linguistic salience. If the first and the second language share similar linguistic features (i.e., L1-L2 proximity), facilitatory effects are predicted to occur during L2 processing (e.g., Wesche & Paribakht, 2009).

Whether or not L2 processing is shallower than L1 processing is an unresolved question (e.g., Clahsen & Felser, 2006; Witzel et al., 2012). Nevertheless, factors such as proficiency, cognitive processes including memory capacity, etc. have been found to directly affect the parsers' sensitivity to violations of grammatical principles during online sentence processing. Rodriguez (2008) examined L2 learners' working memory capacity and their sensitivity to

morphological anomalies in Spanish sentences. Participants taking part in Rodriguez's research were English L1-Spanish L2 speakers who were tested on Spanish proficiency prior to their participation in the self-paced reading task. The experimental sentences of the study were adapted from Jiang's (2007) study as shown in example (63).

(63) (a) *La perra duermo en el garaje cuando hace frío. **The dog sleep in the garage when it's cold*.

(b) *El cliente del restaurante pido siempre la misma ensalada. **The customer of the restaurant ask always for the same salad*.

Participants also took a working memory (WM) test. Findings from Rodriguez's experiment showed that there is an association between participants' L1 background and their sensitivity to L2 morphological markers. Rodriguez confirmed that not all L2 learners are insensitive to morphological anomalies, and that the ability to detect morphological errors partly depends on the first language of the learners. Regarding the effects of WM, Rodriguez's experiment did not gather sufficient data to confirm the role of WM in L2 sentence processing.

Booth, Clenton, and Van Herwegen (2018) investigated the effects of L1-L2 proximity on L2 processing by recruiting different groups of L2 participants from different language backgrounds: Japanese, and Italic Indo-European. English native speakers participating in the study as a control group. Participants from the three groups were asked to complete a vocabulary judgment task and take part in the experiment. Results from the experiment suggested that participants with different L1 profiles varied in task performance. Japanese L1 speakers had lower scores for the semantic judgment task than the Indo-European participants. However, for the syntactic processing task, the Japanese participants showed relatively similar performance compared to the Italic Indo-European speakers. Response times Japanese

participants spent on the judgment tasks were also observed to be slower than their European counterparts. Booth et al. (2018) concluded that in addition to L2 proficiency and other extralinguistic factors, L1-L2 proximity or linguistic closeness between the two languages plays an important role in L2 semantic and syntactic processing.

2.3.3.2. L2 Proficiency

Earlier research on L2 processing mostly recruited high proficient L2 participants (Jackson & van Hell, 2011; Tanner & van Hell, 2014). Results from the recent studies in which L2 participants were less proficient in the second language revealed that L2 proficiency levels differentially affect L2 real time sentence processing (e.g., Dekydtspotter, Schwartz, & Sprouse, 2006; Frenck-Mestre, 1997; Hopp, 2006).

Jackson and van Hell (2011) investigated L2 parsing strategies in the processing of English *wh*-subject extractions and *wh*-object extractions, as shown in example (64).

(64) (a) Who do you think ___ met the tourists in front of the museum? (subject-extraction)

(b) Who do you think the tourists met ___ in front of the museum? (object-extraction)

L2 participants in Jackson and van Hell's (2011) study were Dutch natives. Findings from the experiment showed that L2 speakers with lower proficiency displayed poorer performance on the processing of subject-extraction than object-extraction questions. In contrast, L1 and advanced L2 participants did not significantly differ in their responses to the grammaticality judgment task, or in the processing of the two experimental constructions. The results indicated native-like processing among high proficient L2 speakers, providing evidence that runs counter to the Shallow Structure hypothesis suggested by Clahsen and Felser (2006).

Hopp (2017) also observed native-like processing among high proficient German learners of English in the processing of English *which*-questions. Though it was generally assumed that German speakers of English would demonstrate preferences towards a subject reading over an object reading in English as a result of L1 influence (e.g., Welche Kuh tut die Ziege schubsen? - *Which cow is the goat pushing*? '*Which cow is pushing the goat?* / *Which cow is the goat pushing*?'), results from the eye-tracking experiment showed that L2 participants differed in both grammaticality judgment task and eye-monitoring patterns. Low intermediate L2 speakers exhibited processing costs and insensitivity to inflectional cues, which was demonstrated in their bias towards the subject reading of the experimental *which*-object questions. In contrast, high intermediate L2 participants were able to make use of inflectional cues to guide their syntactic reanalysis and final decisions on the reading of the target question. Advanced L2 speakers, as suggested by preceding research (e.g., Jackson & van Hell, 2012), showcased nativelike performance across experimental tasks. Findings from Hopp's (2017) study provided further evidence for the effects of L2 proficiency in real time sentence processing.

Bel, Sagarra, Cominguez, and Garcia-Alcaraz (2016b) also looked into the effects of proficiency on the application of the Position of the Antecedent hypothesis (PAH) in resolution of Spanish anaphoric pronouns (e.g., *El músico saluda al bombero mientras lleva un violín en la mochila*.). Results from the experiment revealed that L2 participants demonstrated sensitivity to the PAH; nevertheless, only L2 participants with advanced proficiency in the second language displayed native-like processing patterns and were found to be immune from L1 interference in resolution of Spanish anaphors.

To the date of the present study, there have been contrasting findings and proposals regarding L2 late acquirers' online processing of sentences. Advocates of the Shallow Structure hypothesis (Clahsen & Felser, 2006) claimed that L2 learners are not highly sensitive to the abstract and complex hierarchical structures during real time processing, thus depending on lexical and semantic cues to guide interpretation of sentences in the second language. Those standing against the SSH, on the contrary, argued that L1 and L2 processing differences are resulted from other constraints and limitations which are not only restricted to learners' proficiency in the second language (e.g., Jackson & van Hell, 2012).

2.3.3.3. Other Processing Constraints

A prolific amount of empirical research has supported the effects of working memory in L2 sentence processing, particularly in L2 sensitivity to violations of grammatical agreement and deployment of retrieval cues (e.g., Jung, 2018; Kim & Christianson, 2016; Zhou, Rossi, & Chen, 2018). On the contrary, other studies have failed to establish an association between working memory (WM) capacity and participants' performance on processing or production tasks (e.g., Cho, 2018; Rodriguez, 2013). From a more neutral approach, WM capacity was predicted to affect processing only if comprehenders had adequate knowledge of the topic discussed in the text (Joh & Plakans, 2017).

As WM capacity was assumed to mediate between processing costs and task complexity, a number of studies were conducted to measure participants' WM capacity to find the association between working memory and processing (e.g., Cho, 2018). Though Cho's experiment did not reveal any significant indicator of the correlation between WM and task performance, results from the experiment showed that task complexity had negative effects on

L2 participants' accuracy of production tasks (i.e., speaking and writing), while speed and fluency were not affected by the increased task demand.

Contrary to findings from Cho's (2018) research on L2 production, Jung (2018) did not find any significant effects of task complexity on participants' scores of comprehension tasks, but high memory capacity was found to benefit participants in processing and comprehension. Jung (2018) suggested that phonological short-term memory should be taken into consideration in adjacent to complex working memory to unveil the direct effects of WM capacity on L2 processing and comprehension.

As this study mainly focuses on cue-based retrieval mechanisms, it is necessary to review the literature on memory in light of a cue-based framework. The direct-access retrieval mechanism prioritizes a content-addressable principle which suggests that the content of memory representations determines retrieval of cues (Lewis et al., 2006; Martin & McElree, 2016; Engelmann et al., 2019). The content of a memory representation in the direct-access mechanism is weighted in terms of distinctiveness against other representations in memory so that the target representation will eventually be retrieved. However, retrieval is likely to be affected by cue-based interference effects which are resulted from the feature-match between different representations in memory. The overlap in content between representations might lead to cue overload which consequently increases processing difficulty and likelihood of misretrieval (McElree, 2006; Oztekin & McElree, 2007).

Martin (2018) examined the types of cues and the interaction between cues and item features involved in the processing of elliptical sentences (e.g., *Because Jane got the meal that the takeaway sold that night, John* ***did too/*was too****, as usual.*). Results from the two ERP experiments showed that different types of cues were deployed simultaneously during retrieval,

and that processing and comprehension of the experimental materials was determined by the interaction of cues and the content of memory representations. Based on the evidence drawn from the experiments, Martin (2018) suggested a computational mechanism which supports the integration of cues with the addressable contents in memory through a linear (i.e., additive) weighting scheme that takes into account properties of vector addition.

Elaborating on Martin's (2018) findings, Parker (2019) investigated how cues are combined (i.e., linear or nonlinear) into a single retrieval probe that guides retrieval of antecedent in reflexive-antecedent dependencies. Results from Parker's study provided evidence for a nonlinear (i.e., multiplicative) cue combination approach for reflexive licensing. The findings went against earlier models that suggested a linear combination of all cues during retrieval.

The prosody-memory integrated model suggested by Pratt and Fernández (2016) was based on the principle of content-addressable memory retrieval which takes into consideration cue-based interference effects resulted from the interaction between multiple feature-matching contents in memory (Lewis & Vasishth, 2005). Engelmann et al. (2019) developed the original cue-based memory retrieval model (i.e., LV05 model) (Lewis & Vasishth, 2005) into the LV05+IP+MAC model that accounts for item prominence (i.e., IP) and multi-associative cues (i.e., MAC) during retrieval. Compared to the original LV05, the revised model was suggested to better explain the cognitive processes underlying the processing of different syntactic configurations. By testing the predictions of Pratt and Fernández's (2016) suggested model, with consideration given to the revised LV05 model (i.e., LV05+IP+MAC), this present study is expected to shed further light on the computational mechanism of anaphora resolution in general and reflexive-antecedent dependencies in particular.

2.3.4. L2 Processing Hypotheses

The present study looks into the processing and resolution of anaphoric reflexives, taking into consideration the three theoretical accounts: (i) memory retrieval, (ii) implicit prosody, and (iii) good-enough processing and online cognitive equilibrium. Henceforth, this section focuses on the review of the following hypotheses which are suggested to be applicable in both L1 and L2 processing: Shallow Structure hypothesis (Clahsen & Felser, 2006, 2018); Declarative/Procedural Model (Ullman, 2004); Good-Enough Processing hypothesis (Ferreira et al., 2002; Ferreira, 2003; Ferreira & Patson, 2007); and Online Cognitive Equilibrium hypothesis (Karimi & Ferreira, 2016).

2.3.4.1. Shallow Structure Hypothesis

The Shallow Structure hypothesis (SSH) (Clahsen & Felser, 2006) predicted that during sentence processing, L2 learners rely heavily on lexical, semantic, and pragmatic information to steer their decision toward certain interpretation of a sentence. L1 speakers, on the contrary, are more likely to adopt a full processing route that sanctions complex syntactic computation. Clahsen and Felser (2006, p. 32) claimed that "the syntactic representations adult L2 learners compute for comprehension are shallower and less detailed than those of native speakers".

(65) (a) [The nurse] who [the doctor] **argued** [that…

AGENT THEME

(b) [The nurse] who [the doctor] argued [that the rude patient] had **angered**…

THEME

(c) [The nurse] who [the doctor] argued [that the rude patient] had **angered**] is

EXPERIENCER

refusing to work late.

Regarding the resolution of filler-gap dependencies as shown in (65), Clahsen and Felser (2006) proposed that L1 speakers exhibited parsing preferences following the Subjacency Principle. Under this principle, the complementizer *that* is syntactically treated as the beginning of a subordinate clause in adjacent to the clause in which the relative pronoun *who* is used to refer to *the nurse*. In contrast, as a result of excessive reliance on non-structural information, nonnative speakers biased towards immediate attachment of the complementizer *that* to *doctor* and *argued*, indicating their thematic parsing preference over a syntactic parsing preference. L2 speakers were predicted to preferably attach modifiers to the item that bears the lexical information perceived to directly relate to the modifiers. The SSH claims that L2 speakers do not make use of syntactic information as efficiently as L1 speakers, and such shallower syntactic processing is not the result of L1 influence (e.g., Dussias, 2003; Felser et al., 2003; Papadopoulou & Clahsen, 2003).

Clahsen and Felser (2006) also noted that the SSH is applied not only to L2 speakers, but can also be extended to L1 speakers, as supported by findings from earlier research on L1 processing of garden path sentences (e.g., *While Anna dressed the baby played in the crib.*) and other types of syntactic constructions in which lexical-semantic salience was observed to misguide L1 speakers' interpretation of the sentence (e.g., *The dog was bitten by the man*.) (e.g., Ferreira et al., 2002). L1 speakers' shallow syntactic processing accordingly paved the way for the development of the Good-enough hypothesis (Ferreira et al., 2002; Ferreira, 2003; Ferreira & Patson, 2007) and the Online Cognitive Equilibrium hypothesis (Karimi & Ferreira, 2016). The hypotheses suggest that under high cognitive task demand, comprehenders tend to favor a simple, heuristic analysis over a detailed syntactic parsing of the sentence.

Over the course of ten years since the SSH was first introduced, a number of studies have been conducted and claimed to provide counterevidence to the hypothesis (e.g., Witzel et al., 2012). Opponents to the hypothesis critically argued against the proposal that L2 speakers have shallower processing due to deficiency in grammatical knowledge of the second language, and criticized the hypothesis for not providing sound evidence to the shallow processing that has also been observed in native speakers. In response to the counter-arguments made towards the hypothesis, Clahsen and Felser (2018) revisited the SSH, presenting a number of short-comings in the studies that criticized the SSH, and providing explanations and arguments for the validity of the SSH. Clahsen and Felser (2018) proposed that the SSH should be treated as a multiple-pathway hypothesis which is not supposed to be specifically assigned to L2 speakers, but also extended to bilinguals and native speakers. A promising way to refine the SSH is to take into consideration the time course and the weighting of different constraints in sentence processing by a wide and diverse range of language speakers who differ also in their age of acquisition.

2.3.4.2. Declarative-Procedural Model (DP Model)

The DP Model (Ullman, 2001) suggests that there are two systems of learning and memory in the human brain that are responsible for the acquisition of a language: *declarative* and *procedural* memory. The former one has been attributed to the storage of explicit (conscious) knowledge such as the lexical-semantic knowledge of a language, and the events/incidents that a person has experienced in life. Recently, declarative memory has been extended to the inclusion of different types of implicit knowledge alongside knowledge of explicit facts and events (Chun, 2000; Henke, 2010; Ullman & Pullman, 2015). Procedural memory, on the other hand, refers to the learning of implicit knowledge including perceptual-

motor and cognitive skills. The system puts emphasis on the necessity of gradual learning and repetition of knowledge exposure which is considered the foundation of successful learning.

With respect to language learning, the declarative and the procedural system of the DP Model respectively correspond with the learning and processing of lexical knowledge (e.g., regular/irregular morphological forms, phonological forms, and semantic components of the word, etc.); grammatical knowledge; and the hierarchical syntactic structure of a language (e.g., grammatical rules). The interaction between the two memory systems in language acquisition is proposed to resemble the interaction in other domains. At the earlier stage of learning, a learner makes use of the declarative memory to store complex morphological forms of words (e.g., *book* → *books*, *play* → *played*) in memory, concurrently acquiring the underlying syntactic structure and grammatical rules of the language, and gradually storing such knowledge in the procedural memory. As a result, through constant exposure to the language, procedural-based knowledge may override declarative-based one, thus increasing the likelihood of the learner to develop grammatical automaticity in language comprehension and production.

Regarding L1 and L2 processing, Ullman (2001, 2016) stated that L1 and L2 speakers differ in their dependence on the two types of memory systems. While L2 relies more strongly on declarative memory for grammatical knowledge processing, L1 speakers, on the contrary, store the implicit linguistic knowledge of a language in their procedural memory. Though Ullman (2016) predicted that L2 learners "may never proceduralize their grammar to the same extent as L1 learners" (p. 958), the researcher also claimed that there are a variety of factors that may affect L2 learners' proceduration of a language's grammar. L2 learners who have better procedural memory will have more profound grammatical knowledge. High proficiency is also

suggested to indicate good knowledge of grammar, irrespective of the learners' reliance on either procedural or declarative memory.

2.3.4.3. Good-enough and Online Cognitive Equilibrium Hypotheses

Under the Good-enough Processing hypothesis (GE hypothesis) (Ferreira et al., 2002; Ferreira, 2003; Ferreira & Patson, 2007), task demands determine the extent to which language users make use of all of the available linguistic information during processing. A heuristic approach to processing saves cognitive effort by applying quick and simple interpretation of the input. However, such a simple and heuristic processing strategy may lead to a higher probability of misinterpreting the sentence. Evidence that demonstrates the deployment of underspecified linguistic representations following a heuristic processing strategy is manifested in the processing of garden path sentences, as shown in example (66).

(66) While Mary bathed the baby played in the crib.

This garden path construction induces processing difficulty as *the baby* is likely to be mistakenly assigned a syntactic function of an object to the verb *bathed*. The remaining part of the sentence soon disregards this interpretation, resulting in the reanalysis of the sentence and the reassignment of the syntactic role to the noun *the baby*. Responses that comprehenders gave to the comprehension question *Did Mary bathe the baby?*, and their reconstruction of the sentence illustrated that comprehenders do not always compute accurate syntactic structure of sentences during processing (Christianson et al., 2001; Ferreira et al., 2001).

(67) (a) *The dog bit the man*. (active voice/plausible)

(b) *The man bit the dog.* (active voice/implausible)

(c) *The man was bitten by the dog.* (passive voice/plausible)

(d) *The dog was bitten by the man.* (passive voice/implausible)

Further evidence pointing to the GE hypothesis emerged from the processing of sentences such as (67) (Ferreira, 2003). Participants in Ferreira's (2003) study were asked to identify the thematic roles (i.e., *agent* or *patient*) taking part in the event described in the experimental sentence. Findings from the experiment showed that the passive voice/implausible condition (67d) produced the highest rate of erroneous assignment of thematic roles, which was an indicator of how participants' heuristic processing gravitates the thematic structure of a sentence (Agent - Verb - Patient) towards the Noun - Verb - Noun syntactic hierarchy. It was proposed that heuristic processing does not necessarily condemn algorithmic processing, but rather be activated simultaneously with or prior to algorithmic processing during sentence processing (Karimi & Ferreira, 2016).

Though evidence has been generated to support a heuristically good-enough processing approach at both inter- and intra-sentential levels (e.g., Christianson, 2016; Dwivedi, 2013; Klin, Weingartner, Guzman, & Levine, 2006), the underlying determinant that directly affects how deep or shallow language processing is still remains an open question. Upon the initial encounter of a linguistic input, comprehenders immediately start to process the available information, which might disturb the equilibrium state of the cognitive system (Karimi & Ferreira, 2016). Karimi and Ferreira's (2006) Online Cognitive Equilibrium hypothesis (OCE) is based on the original hypothesis suggested by Piaget (1952, 1977) which centers around human cognitive development and their adaptation to the newly emerged environment. There are two processes that are assumed to be involved in human cognitive adaptation: *assimilation* concerning integration of information from the new environment into the existing, already established cognitive system; and *accommodation* referring to the changes that are forced to occur in the existing cognitive structure. Cognitive equilibrium is established once there is a

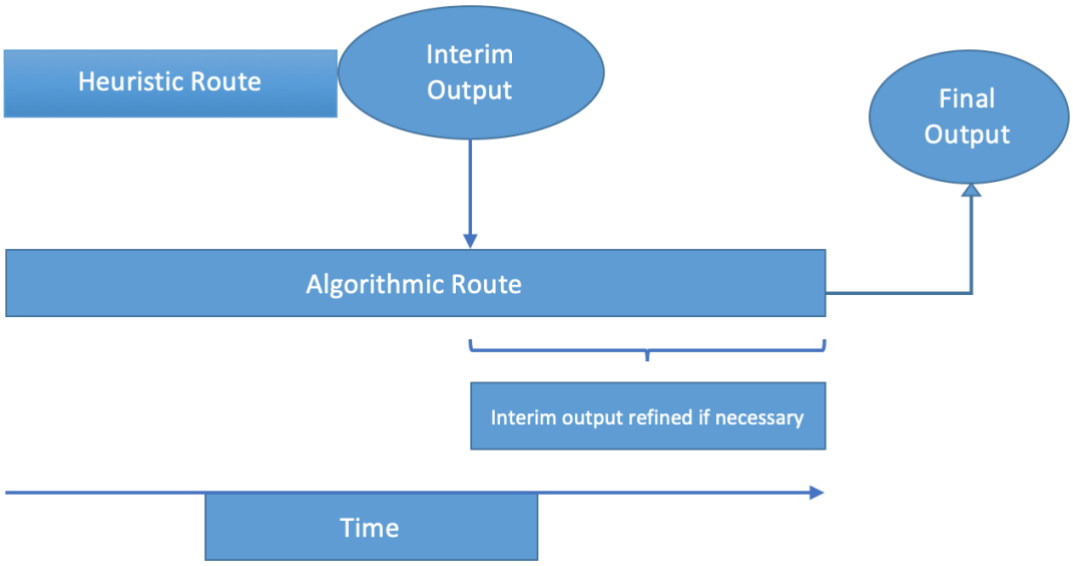
balance between assimilation and accommodation. In contrast, whenever there is an unbalance between the two (i.e., information from the new environment does not match information stored in the existing cognitive system), disequilibrium arises.

Mapping the original cognitive equilibrium hypothesis (Piaget, 1952; 1977) into the cognitive schemata required for language processing, Karimi and Ferreira (2006) suggested that comprehenders' cognitive equilibrium is disturbed as soon as a sentence or an utterance is presented, giving rise to the state of cognitive disequilibrium. During the course of processing, comprehenders gradually move away from the cognitive disequilibrium to eventually get back to the equilibrium state. Language processing, under Karimi and Ferreira's (2006) Online Cognitive Equilibrium hypothesis, is proposed to be responsive to the changes in the equilibrium states. Correspondingly, Karimi and Ferreira offered two principles underlying the OCE Hypothesis, shedding light on comprehenders' sensitivity to such changes. The first principle concerns how the human cognitive system tries to reach the equilibrium state at the earliest opportunity during processing. The second principle hypothesizes that once reaching the equilibrium state, the cognitive system is likely to remain in such a state until there is a reason that is strong enough to force the system to move out of equilibrium.

The OCE hypothesis was suggested to provide the mechanism underlying the GE hypothesis on the account that heuristic processing which is realized by the application of simple rules rather than complex algorithmic computation puts the cognitive system in the equilibrium state. Once reaching equilibrium, the system prefers staying in the state for as long as possible, resulting in the abandonment of detailed syntactic processing.

Figure 3

Model of Language Processing – OCE Hypothesis (Karimi & Ferreira, 2016)



Karimi and Ferreira (2016) tested the two principles of the OCE hypothesis in the processing and resolution of referents. The experimental materials were manipulated in terms of syntactic complexity and linear distance between the two antecedents and the pronoun, as illustrated in (68).

(68) (a) The wizard disagreed with the knight (*short–short*)

(b) The wizard who was confused and depressed by the irreparable situation disagreed with the knight. (*long–short*)

(c) The wizard disagreed with the knight who was confused and depressed by the irreparable situation. (*short–long*)

→ He suddenly came up with a good idea to solve the problem.

Visual displays corresponding to the sentences were followed by comprehension questions for certain experimental sentences. Results from the eye-monitoring experiment

showed that the resolution of the pronoun was faster when the complexity of the sentence increased, which can be explained in light of the two principles of the OCE hypothesis.

Pratt and Fernández (2016) suggested a processing model which was based on the framework of the Good-enough hypothesis. Under Pratt and Fernández's model, language users are expected to opt for a heuristic approach that favors general comprehension over detailed syntactic processing, especially under heavy cognitive load. The predictions of the model also correspond to the two principles of the OCE hypothesis.

2.3.5. L2 Reading

Reading in the second language is considered a cognitive challenge to L2 readers as nonnative speakers have to cognitively process sentences in a language that they have not fully acquired. Earlier research on the neurocognitive processes involved in L2 reading cast light on an even bigger question: whether the psycho-cognitive processes in L2 reading resemble those in L1 reading, or there are unique, distinctive neurocognitive processes underlying L2 processing that are not experienced by L1 speakers.

2.3.5.1. L1 Transfer in L2 Reading

Cummins (1979) proposed the Linguistic Interdependence hypothesis, suggesting that L1 knowledge can be instrumental in the acquisition of L2. During L2 acquisition, certain knowledge of a learner's native language can be transferred into the learning of L2 to facilitate L2 acquisition. However, as Clark (1979) enlarged on this idea with his Linguistic Threshold hypothesis, L1 positive transfer can only occur during L2 learning on the condition that the L2 learner has already obtained certain linguistic knowledge in the second language. In other words, L2 proficiency determines the extent to which L1 knowledge can be transferred into L2 to facilitate L2 acquisition. Empirical research has provided evidence to support the validity of

both hypotheses, indicating that both L1 reading skills (e.g., Droop & Verhoeven, 2003; Van Gelderen, Schoonen, Stoel, de Glopper & Hulstijn, 2007; Verhoeven, 1991, 1994, 2000) and L2 proficiency (e.g., (e.g., Bernhardt & Kamil, 1995; Schoonen, Hulstijn, & Bossers, 1998) have certain effects on the learning of the second language in general and L2 reading in particular.

To test the interplay between the Linguistic Interdependence hypothesis (Cummins, 1978) and the Linguistic Threshold hypothesis (Clark, 1979), Jiang (2011) recruited 246 EFL Chinese students to participate in the study. Participants' scores in the admission exams to college were used as measures of L1 literacy and L2 proficiency. Participants were required to take an L2 reading comprehension test adapted from TOEFL[2] and another reading test designed by the researcher. Findings from the study showed that L1 literacy was relatively correlated with L2 proficiency, and L2 proficiency was found to be congruent with participants' performance in the reading tests. Nevertheless, Jiang did not observe any significant correlation between advanced L2 proficiency level and higher rate of L1 positive transfer into L2 reading. It was assumed that the Chinese participants in the study were not proficient enough in the L2 for the L1 to be successfully transferred into L2 learning. Results from Jiang's research presented an avenue for further studies on the relationship between the Linguistic Interdependence and the Linguistic Threshold hypotheses accounting for different populations of L2 speakers with varied L2 proficiency levels.

Over the past decades, a wide range of studies have generated contradictory findings regarding the Linguistic Threshold and Linguistic Interdependence hypotheses, putting forward an underlying determinant of L1 transfer in L2 learning and acquisition (e.g., August, 2006; Carrell, 1991; Park & Chae, 2000). Park (2013) revisited the Linguistic Threshold hypothesis in

[2] Test of English as a foreign language

L2 reading and made attempts to test the predictions of the hypothesis by examining the interaction between L1 reading, L2 knowledge and L2 reading comprehension. Participants of the study were 2666 Korean EFL learners who were applicants to university in Korea. Participants were measured on their L1 literacy, L2 knowledge, and L2 reading comprehension. Findings from the study did not support the Linguistic Threshold hypothesis as there were other observed factors affecting the threshold level such as L1-L2 proximity, L2 proficiency, task complexity, context of learning, etc. In contrast, regarding L2 reading comprehension, L1 reading and L2 knowledge were found to play a significant role in participants' performance in the reading tests, suggesting pedagogical implications in the teaching of L2 reading skills to EFL learners.

L1 influence in L2 reading has also been studied in light of neurocognitive science (e.g., Jeong, Sugiura, Sassa, Haji, Usui, Taira, Kawashima, 2007; Kim, Qi, Feng, Ding, Liu, & Cao, 2016; Yokoyama, Kim, Uchida, Miyamoto, Yoshimoto, & Kawashima, 2013). Results from these studies indicated that L1-L2 similarities in syntactic features correlate with activation patterns of the brain. Kim, Liu, & Cao (2017) investigated the brain activity on a visual word rhyming judgment task. Participants of the study were English natives, and L2 speakers of English whose first languages are Korean and Chinese. It was expected that the two groups of L2 speakers would engage in distinctive neurocognitive processes to complete the tasks. Findings from the study showed that the two bilingual groups predominantly relied on their L1 reading to perform L2 reading tasks. Greater involvement of L1 in L2 reading was also observed in L2 speakers who were more proficient in L2. Kim et al.'s (2017) research provided further evidence in neuroimaging to support the proposal that the brain regions required in L1 reading also function in L2 reading and that the extent of L1 transfer was found to be correlated

with L2 proficiency levels. Results from Kim et al.'s brain imaging study further supported the Linguistic Threshold hypothesis, such that it is necessary for L2 speakers to reach a certain proficiency level, i.e., a threshold, for their L1 reading to be mapped onto L2 reading.

2.3.5.2. L2 Automaticity in Reading

L2 reading fluency partly depends on L2 learners' lexical automaticity (i.e., the ability to automatically access the accumulated collection of known words in the second language) (Schmitt, 2010). Lexical automaticity is a reliable indicator of the development of reading fluency. However, that an L2 learner knows a word in L2 does not warrant his/her automaticity in using or recalling the word when required by the context. Automaticity is considered the ultimate result of automatization which is reflected in a learner's development of his/her L2 implicit knowledge obtained through constant exposure to L2 or frequent L2 practice (DeKeyser, 2001). Automaticity including lexical automaticity is defined as a fast, non-stop, effortless, and unconscious process (Segalowitz & Hulstijn, 2005) which has been suggested to facilitate L2 reading because automaticity releases L2 readers from disruptions caused by distractors. In addition, lexical automaticity also reduces cognitive effort required at the word level so that the cognitive system could save up resources for other task demands that require more complex syntactic computation, especially during online sentence processing.

However, whether automaticity at low levels (e.g., lexical level) guarantees overall improvement in reading comprehension still remains an unresolved question. Fukkink, Hulstijn, & Simis (2005); and Perfetti (1985, 1988) consistently found that training in L2 word recognition and automaticity at the word level does not always guarantee improvement in L2 reading comprehension. Alternatively, post-lexical automaticity was proposed to have a more direct influence on reading comprehension. Grabe and Stroller (2002, p. 37) suggested that "L2

students need some foundation of structural knowledge and text organization in the L2 for more effective reading comprehension". This suggestion puts emphasis on the importance of text segmentation into meaningful units, and the post-lexical automaticity at the syntactic level to promote the improvement in L2 reading fluency and comprehension. Similar findings regarding the contribution of prosody over lexical automaticity to reading comprehension were also found in L1 silent reading comprehension. Paige, Raskinski, Magpuri-Lavell, and Smith (2014) did not observe any positive effects of word recognition automaticity on L1 reading comprehension. On the contrary, Paige et al. (2014) noticed that prosody functions as a liaison between reading automaticity, reading comprehension, and reading fluency. Implicit prosody projected onto reading texts in the form of text segmentation will be discussed in section 2.3.5.3 that follows.

2.3.5.3. Text Segmentation in L2 Reading

The effects of text segmentation on sentence processing and comprehension have been generally supported by empirical research (Hijikata, 2005; Kadota, 1982; Kadota & Tada, 1992; Kadota, Yoshida, & Yoshida, 1999; Yubune, 2012). Kadota (1982) observed that compared to the word-by-word segmentation, the implicit prosody that resembles the syntactic parsing of a sentence was better at facilitating L2 readers with low proficiency in comprehension. Similar results were found in Kadota and Tada's (1992) study in which the experimental stimuli were manipulated across the three presentation formats: word-by-word, phrase-segment, and whole-sentence. Findings from the experiment revealed that the phrase-segment format was most facilitative for participants during processing. Word-by-word segmentation was found to be disruptive, while the other two formats: phrase-segment and clause-by-clause were observed to assist L2 participants, reflected in increased accuracy in the comprehension task and reduced total reading time in the reading tasks (Kadota et al., 1999). Findings from Kadota's (1982) and

Kadota and Tada's (1992) research are not completely in line with findings from some earlier research with respect to the effects of word-by-word parsing on reading (e.g., Cromer, 1970).

One of the earliest studies on the effects of text segmentations in reading comprehension was conducted by Cromer (1970). Cromer investigated the text segmentation strategies applied by readers (i.e., college students) differing in reading skills. Participants were put into four groups conforming to their reading skills. The participants were asked to read the experimental sentences presented in the following formats: *whole sentence*, *word-by-word, phrase segments*, and *random word groupings*. Findings from the experiment were in contrast with Cromer's initial hypotheses, such that not all of the groups benefited from the phrase-segment format which is the most expected format. The group of least skilled readers found the word-by-word non-disruptive, which also contradicted Cromer's initial expectations. In fact, participants with poorest reading skills found the word-by-word segmentation format most beneficial during processing. The whole-sentence and random word groupings were not found to be as disruptive or disturbing as Cromer had predicted them to be. Cromer concluded that phrase segmentation did not have any facilitatory effects on readers' comprehension in L2 silent reading if the comprehenders have not yet acquired certain vocabulary skills in the second language.

Yamashita and Ichikawa (2010) extended Cromer's (1970) research to shed light on the association between reading fluency and chunking strategies. Using self-paced reading, Yamashita and Ichikawa presented the experimental materials in four segmentation paradigms: *whole text*, *single words*, *meaningful chunks*, and *fragmented word groups.* Participants of the study were 48 Japanese college students who had intermediate to advanced proficiency levels of English. Each reading text was followed by five comprehension questions. The samples of the *meaningful chunks* and *fragmented word groups* are presented in (69).

(69) (a) ***Chunked***: The origin of Australian Rules Football / is unclear. / Some people say / it might have developed / from an ancient game / in which a ball made of kangaroo skin / was kicked around.

(b) ***Fragmented***: The origin of Australian / Rules Football is unclear. / Some people say it might / have developed from an / ancient game in / which a ball made of kangaroo / skin was kicked around.

Yamashita and Ichikawa expected that the *meaningful chunks* format corresponding with the syntactic structure of the sentence would facilitate participants with lower proficiency level, while the *fragmented word groups* would hinder processing, thus negatively affecting comprehension. Participants with advanced proficiency level, on the contrary, were hypothesized not to be affected by the manipulations of text presentation formats. Results from the experiment demonstrated that the *meaningful chunks* format did not significantly facilitate any groups of participants in reading comprehension, while the *fragmented word groups* format took a heavier toll on comprehension in the low proficiency group as compared to the higher proficiency. Though the *single words* format cost the advanced participants more time on reading, their performance in comprehension tasks were not negatively affected.

Pratt and Fernández (2016) also observed different effects of implicit prosody on different groups of participants during sentence processing. While the phrase-segment presentation format facilitated L1 participants in syntactic computation and reading comprehension, the word-by-word format had an adverse effect on comprehension accuracy. L2 participants, in contrast, were not observed to benefit from the phrase-segment format as much as the L1 group. L2ers' performance on grammaticality judgment and their responses to comprehension questions was even found to be hammered by the phrase-segment presentation.

Results from Yamashita and Ichikawa's (2010) and Pratt and Fernández's (2016) studies indicate that the effects of text segmentation vary across different groups of participants who differ in L2 proficiency. Low proficient readers who are used to reading texts at the space of word by word may find the corresponding text segmentation most beneficial in reading. Whereas, speakers with higher proficiency levels who are able to syntactically parse sentences into meaningful chunks may be facilitated by the phrase-segment format. Advanced L2 participants with nearly native-like proficiency may withstand disruptions from the fragmented text presentation, thus managing to maintain good performance across experimental tasks.

This Chapter has reviewed the followings: (i) fundamentals of anaphora resolution (syntactic constraints, non-syntactic constraints, and cue-based memory retrieval) with a focus on the resolution of reflexive-antecedent dependencies; (ii) the interplay between prosody and sentence processing with respect to syntax and working memory; and (iii) second language processing including factors that might affect processing in the second/foreign language. In this Chapter, I have highlighted the three areas of inquiry that provide the standing ground for the present study: (i) Cue-based Memory Retrieval, (ii) Implicit Prosody, and (iii) Online Cognitive Equilibrium, and Good-enough Processing hypotheses. The following Chapter will describe in detail the experimental methods and data collection procedures of the study.

Chapter 3. Experimental Investigation

The purpose of the present study is to investigate L1 and L2 processing and resolution of reflexive-antecedent dependencies in light of the processing model suggested by Pratt and Fernández (2016) that integrates implicit prosody (Bader, 1998; Fodor, 1998, 2002) and cue-based memory retrieval (Lewis & Vasishth, 2005; Engelmann et al., 2019) into the framework of Good-enough Processing (Ferreira et al., 2002; Ferreira, 2003; Ferreira & Patson, 2007) and Online Cognitive Equilibrium hypotheses (Karimi & Ferreira, 2016). It is expected that a similar pattern of results would be observed in the present study, such that (i) manipulations of syntactic complexity and implicit prosody would affect L1 and L2 processing of reflexive anaphors, (ii) different groups of participants would be differentially affected by text presentation formats (i.e., implicit prosody) during processing, and (iii) language proficiency would function as a mediator between implicit prosody, memory access, and reading comprehension.

3.1. Methods

This section presents the methods for data collection including descriptions of participants, recruitment procedures, and justifications for research design (i.e., design of experiment and supplementary tasks).

3.1.1. Participants

As the present study aims to test L1 and L2 processing of reflexive-antecedent dependencies, two groups of participants: (i) English natives, and (ii) Vietnamese L1-English L2 speakers were recruited for the study. There were 74 participants in total. None of the participants had any vision problems. Initially, 81 participants took part in the study. However, seven of them were then excluded from data analysis for either being bilingual since birth (as

self-declared in the pre-experiment questionnaire, *n = 3*) or failing to follow instructions during experiment (*n = 4*).

L1 participants: Thirty-two English native speakers (twenty-one of them were British, the rest were American, Australian, Irish, and South African) took part in the study via public recruitment posts on the Facebook group for Expats in Hanoi, Vietnam. Their age ranged from 22 to 37 (*Mean = 27.5, SD = 3.9*). The participants were compensated 500,000 VND (about 20 USD) for their participation in the study.

L2 participants: Forty-two Vietnamese-L1 English-L2 speakers were recruited via emails and in-person contacts. Twenty of the L2 participants were English language teachers at the English Department, Hanoi University in Hanoi, Vietnam. The rest were students majoring in English language studies in the same Department. The age range of the L2 group was between 19 and 39 ($Mean_{Age} = 25.1, SD_{Age} = 5.2; Mean_{Age\ of\ L2\ acquisition\ –\ AoA} = 8.6, SD_{AoA} = 2.3$). Participants were pseudo-randomly assigned to a reading format and a stimulus block. All participants signed the consent form before the experiment to indicate their voluntary participation in the study.

3.1.2. Experimental Design and Materials

The experimental design was 2x2x3 and manipulated in (i) syntactic structure (i.e., *simple, complex*), (ii) grammaticality (*target-match, target-mismatch*), and (iii) implicit prosody in the form of text presentation formats (*word-by-word*, *phrase-segment*, *whole-sentence*).

The stimuli were arranged in a 2x2 factorial design crossed two factors: syntactic structure (*simple, complex*) and grammaticality (*target-match, target-mismatch*) (see Table 2). Experimental stimuli were adapted from Patil et al.'s (2016) research with the *match-interference* and *mismatch-interference* conditions respectively corresponding to the *target-*

match (i.e., *grammatical*) and *target-mismatch* (i.e., *ungrammatical*) conditions of the present study. The accessible antecedent (e.g., *the actress*) and the distractor(s) (e.g., *Mary/John*) were both assigned the syntactic role of a subject to increase the strength of feature match between the two antecedents, which was expected to increase cognitive load and interference effects during retrieval (Patil et al., 2016; Van Dyke & McElree, 2011). Manipulation of syntactic complexity was manifested in the inclusion of a past participle reduced relative clause in the complex configuration in place of the adverbial phrase in the simple configuration. The experimental sentences (48 experimental items and 112 fillers) were matched in length, i.e., each contained 19 words. The vocabulary items used in the experimental sentences were selected from the 10,000 most frequently used words in the iWeb corpus (English Corpora, 2019) to minimize processing difficulties due to unfamiliar vocabulary.

For the construction of experimental stimuli, the study chose not to use stereotypical gender nouns (e.g., *engineer, nurse*), but instead, opted for the use of definitional gender nouns (e.g., *actress, king*). There are two motivating reasons for the utility of definitional gender nouns in the experimental sentences: first, to maximize the feature match/mismatch between retrieval cues and item features (e.g., *actress, Mary, herself*); second, to eliminate the gender-bias caused by stereotypical gender nouns since the gender in stereotypical gender nouns (e.g., *doctor, teacher*) is often subject to different interpretations due to diverse personal beliefs and social perceptions of gender (Canal, Garnham, & Oakhill, 2015).

A sample set of experimental items is given in Table 2.

Table 2

Sample Set of Experimental Items

Syntactic structure	Grammaticality	Sentence
Simple	**Grammatical** (target-match)	a. The actress that Mary interviewed at the awards ceremony about two years ago described herself as an extreme workaholic.
	Ungrammatical (target-mismatch)	b. The actress that John interviewed at the awards ceremony about two years ago described himself as an extreme workaholic.
Complex	**Grammatical** (target-match)	c. The actress that Mary interviewed at the awards ceremony held outside the theater described herself as an extreme workaholic.
	Ungrammatical (target-mismatch)	d. The actress that John interviewed at the awards ceremony held outside the theater described himself as an extreme workaholic.

Forty-eight experimental items and 112 fillers were distributed across four counter-balanced lists following Latin Square design. Each list was categorized into four blocks, generating 16 blocks in total (each block contained 40 sentences). Half of the experimental sentences, including the fillers, were ungrammatical. The fillers were constructed conforming to various syntactic configurations: possessive adjective-embedded clauses, PP-embedded subjects, complement clauses, adverbial clauses, relative clauses, and clauses that contained countable/uncountable nouns (see Appendix D for a complete list of the fillers).

3.2. Data Collection

3.2.1. Norming Procedures

All of the experimental materials (fillers included) in the grammatical condition (96 experimental sentences and 56 fillers) were normed by native English speakers. The two

norming surveys were conducted online via Google Forms (Google LLC, 2019). All participants were asked for their consent before proceeding to the norming survey. For the first survey, eighteen American-English speakers were recruited via word of mouth. Participants were asked to judge the plausibility of the experimental sentences (i.e., the likelihood of hearing the sentences in natural discourse) on a scale from 1 to 5 in which 1 indicates *implausible*, 2 = *somewhat implausible*, 3 = *plausible*, 4 = *highly plausible*, and 5 = *totally plausible*. Ratings were averaged, and the sentences with a mean score below 3.0 were revised and normed the second time.

The second norming survey consisted of 26 revised experimental sentences from the first survey. The second survey was distributed to 29 students at Ohio State University whose first language was English. The students were given bonus credits for their participation in the survey. The plausibility rating scale used in the second norming survey was the same as that of the first survey. Responses were collected within a week, and all of the sentences came back with an average rating of 3.0 and above, indicating that no further revision or norming would be necessary.

3.2.2. Design of Experiment

The 16 blocks of experimental stimuli were pseudorandomized into the three text presentation formats: *word-by-word* (rapid serial visual presentation - RSVP), *phrase-segment* (self-paced reading), and *whole-sentence* (self-paced reading). The experiments were run on the software PsychoPy3 (Peirce, Gray, Simpson, MacAskill, Höchenberger, Sogo, Kastman, & Lindeløv, 2019), version 3.1.2.

Table 3

Reading Presentation Formats

Presentation format	Presentation rate
Word-by-word	RSVP
Phrase-segment	Self-paced
Whole-sentence	Self-paced

The reading experiment started with a greeting followed by instructions on how to proceed throughout the experiment. Participants were first prompted to the four practice items before advancing to the experimental stimuli. Multiple trials of the practice items were allowed until participants felt comfortable to proceed to the actual experiment. Each experiment session lasted for about 40 to 45 minutes, including a reading comprehension test administered upon participants' completion of the reading experiment. Description of the real-time reading test will be provided toward the end of this section.

Each experimental stimulus was preceded by a fixation cross in the center of the screen (with a final time-out of 2000 milliseconds). The fixation cross helped divert the participants' attention to where the first word (or the first phrase) would come up on the screen.

In the *word-by-word* presentation format, each word was set to automatically appear in the center of the screen after every 500 milliseconds. As reading rates fluctuate depending on the reading tasks involved (Carver, 1992, as cited in Primativo, Spinelli, Zoccolotti, Luca, & Martelli, 2016, p. 2), readers might accelerate in their reading speed (600 words per minute) in tasks that require scanning through the text to find the target word. However, they might also slow down in tasks that consume memory (Carver, 1992). Increased cognitive load and memory decay are expected to occur during processing, especially in cases of lengthy and syntactically complex sentences; thus, the presentation rate of 500 milliseconds per word (equivalent to 120

words per minute) was considered an appropriate fixed rate for the *word-by-word* (RSVP) experiment.

The *phrase-segment* and the *whole-sentence* presentation paradigms were presented as self-paced reading experiments. In the *phrase-segment* format, an experimental stimulus was divided into four segmental units conjuring to the syntactic parsing of the sentence:

(70) a. The actress | that Mary interviewed at the awards ceremony | about two years ago | described herself as an extreme workaholic.

b. The actress | that John interviewed at the awards ceremony | held outside the theater | described himself as an extreme workaholic.

The sentence segments appeared centrally on the screen in sequence. The first segment came up on the screen following the fixation cross. The final timeout for each segment was set at 5000 milliseconds. Participants were instructed to pace their own reading, and press the spacebar on the keyboard to advance to the next segment(s) of the sentence as soon as they finished reading the previous one(s). The *phrase-segment* format was predicted to ease processing by reducing cognitive load. Segmentation of sentences into chunks that correspond with the syntactic parsing of a sentence might preserve memory resources, which accordingly prevents cue decay (Pratt & Fernández, 2016). However, since participants could not go back to the previous segment(s), misretrieval was also expected to occur. Further discussion on the probability of misretrieval will be provided in the Discussion chapter.

The reading rate for the *whole-sentence* paradigm was self-paced with the final timeout set at 19000 milliseconds from the initial onset. Participants were encouraged to advance to the subsequent task/trial as soon as they finished reading each sentence. The *whole-sentence* reading format was also expected to facilitate processing since the presentation of the whole sentence on

the screen resembles the format of normal reading in which regressions (i.e., backward gaze moves) are allowed. For the *whole-sentence* format, past information, especially the information that was not "clearly perceived or understood" (Rayner, 2009, as cited in Benedetto, Carbone, Pedrotti, Fevre, Bey, & Baccino, 2015, p. 353), could be re-examined through regressions. As a result, more accurate syntactic analysis and better comprehension were expected in this format of reading.

On each trial, there were two tasks following the presentation of the experimental stimulus. The first task was a grammaticality judgment task that required participants to rate the grammaticality of the sentence on a scale from 1 to 5 (1 = *totally ungrammatical*, 5 = *absolutely grammatical*). Participants were instructed to press either one of the five keys (1 to 5) on the keyboard to indicate their response to the judgment task. To minimize the chance that participants would provide low ratings due to overall dislike of lengthy and complex sentences, participants were encouraged to rate a sentence with respect to how grammatical they thought the sentence was, regardless of their personal impression on the construction of the sentence. The time out of the grammaticality rating task was set at 10000 milliseconds.

The second task was a comprehension task that asked participants to respond to a question that targeted either the general comprehension of the sentence, or the licensing of the reflexive. These two types of comprehension questions were distributed relatively evenly across experimental stimuli. Participants were given a maximum of 8000 milliseconds to respond to the question probe by pressing either the Y key (i.e., for *Yes*) or N key (i.e., for *No*) on the keyboard. Examples of the comprehension probes are given in (71).

(71) The actress that Mary interviewed at the awards ceremony about two years ago described herself as an extreme workaholic.

a. *General-targeted question*: Was the awards ceremony organized three years ago?

b. *Reflexive-targeted question*: Did the actress describe herself as a workaholic?

Participants' grammaticality ratings, their responses to the comprehension questions, as well as their reading times and reaction times (in milliseconds) in every experimental task were then analyzed using mixed-effects models (Baayen, 2008). Details on the selection of mixed models are given in section 4.1.

3.2.3. Data Collection Procedures

Prior to the experiment, L1 participants were asked to complete the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian, Blumenfeld, & Kaushanskaya, 2007); L2 participants completed the Language Profile Questionnaire adapted from the LEAP-Q (see Appendix A and Appendix B). The self-rated reading proficiency questionnaires were used as a reference to the reading test administered after the reading experiment. A correlation test between the self-rated proficiency and the results of the reading test was conducted to get insight into the association between the two measures of reading proficiency. Results and discussion of the correlation test will be provided in the Results and Discussion chapters.

Following the reading experiment, participants were asked to proceed with the reading test. The test was adapted from Reading Test 1, Part 2 in the Cambridge Certificate in Advanced English tests (Vol. 6, Cambridge Books for Cambridge Exams, 2005). Since this study focuses on real time processing that takes into account real time deployment of retrieval cues, the reading test was revised to mimic a real time processing task. The reading passage contained

306 words with 15 gaps. As participants silently read along the passage, they were asked to simultaneously fill in each gap with one word. Responses from the participants were scored on a scale from 1 to 15.

Chapter 4. Results

4.1. Overview of Data Analysis and Statistical Procedures

As the study primarily looks into the effects of implicit prosody during processing of reflexive anaphors with respect to cue-based memory retrieval mechanism, reaction times and responses to experimental tasks were treated as response variables. Groups of participants (*L1, L2*), measures of English reading proficiency, and crossed factors of experimental design (i.e., *syntactic structure*, *grammaticality*, and *text presentation formats*) were analyzed as predictor variables. Mixed-effects models (Baayen, 2008) were used for data analysis. The models were fit to the data with effects of predictor variables as well as all possible interactions between variables justified by the research design.

The two-level factors (i.e., *groups, syntactic structure, grammaticality*) and the three-level factor (i.e., *text presentation formats*) were contrast-coded (i.e., deviation coded) before analysis to extract the main effects from the model outputs (Barr, 2020; Winter, 2019). The continuous variable (i.e., *measures of English reading proficiency*) was centered for a more accurate interpretation of coefficient outputs (Winter, 2019).

Reaction times (RTs) of each phrasal segment (in the *phrase* format), RTs of each sentence (in the *sentence* format), and RTs of responses to the grammaticality judgment and comprehension tasks were log-transformed prior to analysis. Log-transformed RTs that fell beyond the threshold of +/- 3 standard deviations (inclusive of RTs below 250 milliseconds) were trimmed off from the data, resulting in the loss of approximately 0.04% of data points. A summary of cut-off values was given in Table 4.

Table 4

Summary of Cut-off Data

Reaction times	Number of cut-off data points
Segment 1 RT (PHRASE)	1/336
Segment 2 RT (PHRASE)	1/336
Segment 3 RT (PHRASE)	8/336
Segment 4 RT (PHRASE)	2/336
Whole sentence (SENTENCE)	0/264
Grammaticality Rating RT	5/881
Comprehension Question RT	1/873
Total (in percentage)	0.04%

Mixed-effects models were used for data analysis with random intercepts and random slopes for both items and participants (Baayen, 2008; Barr, Levy, Scheepers, & Tily, 2013; Barr, 2013; Barr, 2020; Barr & Debruine, 2021; Bates, Maechler, & Bolker, 2011; Brauer & Curtin, 2017; Meteyard & Davies, 2020; Winter, 2019). A variance-covariance matrix justified by the research design was applied in all models (Barr, 2013; Barr, 2020). When a model failed to converge, selected fixed and random effects were dropped until the model converged again. Selection of random effects to be removed from non-converged models was based on the reasoning that once the random effects for higher-order interactions were still included in the model, removal of random effects for main effects and random effects for lower-order interactions would not inflate Type I error rate (Barr, 2013; Brauer & Curtin, 2017). Likelihood ratio tests (to test the significant effect of particular fixed effects) and restricted maximum likelihood tests (to test the significant effect of random effects and their interactions) were performed to compare the likelihood between models (Winter, 2019).

R software (version 4.0.2) (R Core Team, 2020) was used to run the models and to generate plots and figures. The following R packages (in alphabetical order) were used for data wrangling, analysis, and visualization: *DataExplorer* (Cui, 2020); *extrafont* (Chang, 2014); *lme4* (Bates, Maechler, Bolker, & Walker, 2015); *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2017); *sjPlot* (Lüdecke, 2021); and *tidyverse* (Wickham, Averick, Bryan, Chang, McGowan, François, Grolemund, Hayes, Henry, Hester, Kuhn, Pedersen, Miller, Bache, Müller, Ooms, Robinson, Seidel, Spinu, Takahashi, Vaughan, Wilke, Woo, & Yutani, 2019).

Correlation tests between the self-rated reading questionnaires and the results of the reading test were conducted. The result of the correlation test ($r\ (72) = 0.6, p < 0.001$) indicates a positive correlation between the two proficiency measures.

4.2. Grammaticality Judgment Results

Grammaticality ratings were binary-coded as *accurate* (grammatical sentences: ratings ≥ 3; ungrammatical sentences: ratings ≤ 2) and *inaccurate*. Signal detection theory (Peterson, Birdsall, & Fox, 1954) was applied in the analysis of responses to the grammaticality judgment task to measure participants' ability to distinguish grammatical from ungrammatical constructions as well as their bias in judgment decisions. Specifically, the three following sets of sensitivity measures were calculated for the grammaticality judgment task: (i) overall sensitivity (i.e., *d' value*) and bias (i.e., *c value*) across all participants and items; (ii) overall sensitivity and bias by participant; and (iii) overall sensitivity and bias by item, followed by the analysis of t-tests comparing different sets of d' and c values across groups (*L1*, *L2*), and across syntactic structures (*simple, complex*) (Huang & Ferreira, 2020). Generalized linear mixed-effects models of grammaticality rating accuracy by (i) both groups, (ii) by each of the groups, and (iii) by each

of the text presentation formats were fit to the data. All of the models had random slopes and random intercepts for items and participants.

4.2.1. Overall Sensitivity and Bias

Overall sensitivity (*d' value*) and bias (*c value*) for all items and participants indicated participants' overall ability to differentiate grammatical from ungrammatical sentences. The overall d' values and c values also showed participants' bias in grammaticality ratings. As the 2x2 design of the experimental items crossed the two factors: grammaticality (*target-match, target-mismatch*) and syntactic structure (*simple, complex*), the analysis of the overall sensitivity and bias was conducted in line with each of the syntactic structures. Results of the overall d' values and c values are given in Table 5.

Table 5

Overall Sensitivity and Bias across All Items and Participants

	Simple structure			**Complex structure**		
	Mean	**SE**	**CI**	**Mean**	**SE**	**CI**
Sensitivity (d' value)	0.62	0.10	0.43, 0.82	0.56	0.11	0.35, 0.78
Bias (c value)	-0.63	0.06	-0.75, -0.5	- 0.5	0.06	-0.64, -0.36

The overall values of sensitivity and bias across all items and participants suggested that in general participants could differentiate between grammatical (*target-match*) and ungrammatical (*target-mismatch*) sentences (i.e., confidence intervals did not include the value of 0) in both the simple and complex conditions. Moreover, negative c values indicated participants' bias towards higher ratings (i.e., judging sentences as *grammatical*).

4.2.2. Sensitivity and Bias by Participant

The d' values and c values by participants were calculated over syntactic manipulation: *simple* and *complex*. Each participant had two sets of d' and c values for each of the syntactic

conditions. T-tests were performed to compare the sensitivity and bias for each of the syntactic structures as well as the sensitivity and bias by each of the groups.

Results of the t-test that compared participants' d' values for the simple and complex conditions (*Mean $_{d' \text{ value for simple constructions}}$ = 0.62, SE = 0.1; Mean $_{d' \text{ value for complex}}$ = 0.56, SE = 0.11; t = -0.41; p = 0.68*) revealed that on average, compared to the complex condition, participants had higher sensitivity for grammatical and ungrammatical sentences in the simple condition. However, the difference in participants' sensitivity between the two syntactic conditions was not significant.

L1 and L2 participants did not significantly differ in the ability to distinguish grammatical from ungrammatical constructions even though the native speakers averagely showed higher sensitivity than the nonnative participants (*Mean $_{d' \text{ value of L1}}$ = 0.69, SE = 0.13, CI (0.42, 0.95); Mean $_{d' \text{ value of L2}}$ = 0.52, SE = 0.08, CI (0.36, 0.68); t = 1.07; p = 0.28*). Regarding participants' bias in judgment decisions, compared to L1 participants, the L2 group significantly biased towards judging sentences as *grammatical*. In other words, L2 participants tended to give higher ratings to experimental sentences than L1 participants (*Mean $_{c \text{ value of L1}}$ = -0.4, SE = 0.07, CI (-0.55, -0.24); Mean $_{c \text{ value of L2}}$ = - 0.7, SE = 0.05, CI (-0.8, -0.6); t = 3.11; p = 0.002*).

4.2.3. Sensitivity and Bias by Item

As previously stated in section 3.2, the experimental item of the present study was manipulated in grammaticality (i.e., *target-match, target-mismatch*) and syntactic structure (i.e., *simple, complex*), each experimental item was associated with four conditions as shown in Table 2. The values of sensitivity and bias by item, and the results of the t-tests that compared the d'

values and c values for items in each syntactic condition indicated whether the sensitivity and bias for a target-match (or target-mismatch) sentence in the simple condition was different from that in the complex condition.

Results of the t-tests of sensitivity by item showed that for the grammatical (*target-match*) sentences, participants marginally had higher sensitivity (i.e., better able to distinguish grammatical sentences) in the simple than in the complex condition (*d' value for target-match in simple condition = 1.04, SE = 0.05, CI (0.92, 1.16); d' value for target-match in complex condition = 0.86, SE = 0.07, CI (0.72, 1.00); t = -1.9; p = 0.06*). Conversely, on average, participants did not show sensitivity for ungrammatical (*target-mismatch*) sentences in neither syntactic conditions, which was demonstrated by the negative values of item sensitivity in both of the two conditions (*d' value for target-mismatch in simple condition = -0.32, SE = 0.08, CI (-0.5, -0.14); d' value for target-mismatch in complex condition = -0.24, SE = 0.1, CI (-0.44, -0.04); t = 0.62; p = 0.53*).

The c value by item further revealed that participants biased towards higher ratings for simple items than for complex items (*Mean* $_{c\ value\ for\ simple\ condition}$ *= -0.34, SE = 0.03, CI (-0.4, -0.28); Mean* $_{c\ value\ for\ complex\ condition}$ *= -0.28, SE = 0.03, CI (-0.34, -0.2)*).

4.2.4. Grammaticality Rating Accuracy

Generalized linear mixed-effects models of grammaticality rating accuracy by both groups, by each of the groups, and by each of the text presentation formats were fit to the binary-coded data of rating accuracy. All models had random slopes and random intercepts for items and participants.

The generalized linear mixed model of grammaticality rating accuracy by both groups included fixed effects of *group, format, structure, grammaticality, English reading proficiency,*

interactions between group and format, as well as *interactions between structure and grammaticality*. Results of the model are reported in Table 6.

Table 6

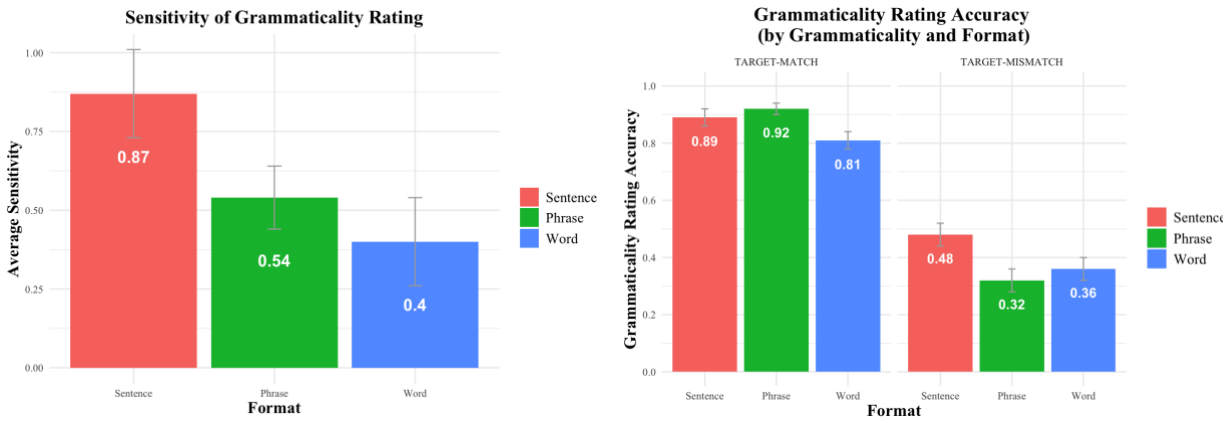
Generalized Linear Mixed-Effects Model of Grammaticality Rating Accuracy by Language Group. Effects of Language Group, Format, Structure, Grammaticality, and English Reading Proficiency

Fixed effects	Estimate	S.E	z value	p value	sig
(Intercept)	0.831	0.118	7.027	0.000	***
Group	0.113	0.256	0.443	0.658	
Word vs. Sentence	-0.549	0.277	-1.979	0.048	*
Phrase vs. Sentence	-0.538	0.267	-2.016	0.044	*
Structure	-0.294	0.202	-1.454	0.146	
Grammaticality	2.720	0.206	13.177	<0.001	***
English reading proficiency	0.120	0.049	2.432	0.015	*
Group x Word vs. Sentence	-0.264	0.575	-0.459	0.646	
Group x Phrase vs. Sentence	-0.219	0.548	-0.400	0.689	
Structure x Grammaticality	-0.902	0.390	-2.314	0.021	*

*RatingAcc ~ Group * Format + Structure * Grammaticality + English reading proficiency + (1 + Structure * Grammaticality | Participant) + (1 + Group * Format * Grammaticality | Item)*

Figure 4

Sensitivity (d' value) of Grammaticality Ratings and Grammaticality Rating Accuracy by Both Language Groups



Results in Table 6 and Figure 4 demonstrated an overall effect of format. Compared with the *word* and *phrase-segment* formats, participants had the highest grammaticality rating accuracy and highest sensitivity in the *sentence* format.

As stated in section 4.2.1, the overall d' value across all items and participants indicated participants' overall ability to differentiate grammatical from ungrammatical sentences in both simple and complex conditions. Results from the model showed that there was an effect of grammaticality, such that grammatical (*target-match*) sentences were rated more accurately than ungrammatical (*target-mismatch*) sentences (see also Figure 4). The higher rating accuracy in target-match sentences was associated with participants' bias towards judging experimental sentences as grammatical (*overall c value = -0.5, SE = 0.06, CI (-0.64, -0.36)*). Furthermore, participants showed sensitivity for grammatical constructions, but not much for ungrammatical constructions in both of the two syntactic conditions, which was reflected in the negative d' values (i.e., below chance) for ungrammatical sentences in both simple and complex conditions: d' value for target-mismatch sentences in simple condition = -0.32, SE = 0.08, CI (-0.5, -0.14); d' value for target-mismatch in complex condition = -0.24, SE = 0.1, CI (-0.44, -0.04).

An additional effect of English reading proficiency suggested an association between reading proficiency and grammaticality rating accuracy. There was also an interaction between structure and grammaticality, denoting that compared to the simple condition, participants rated grammatical sentences in the complex condition less accurately. The findings were consistent with the results of the t-test that compared the d' values for grammatical sentences in the simple and complex conditions ($t = -1.9$, $p = 0.06$), indicating participants' marginally higher sensitivity for grammatical sentences in the simple constructions.

Analysis of grammaticality rating accuracy by each group and by each of the three reading formats was conducted to gain a more profound understanding of the factors that might affect participants' performance in the grammaticality judgment task. Results of the generalized linear mixed-effects models of grammaticality rating accuracy by L1 (with fixed effects of *format, structure, grammaticality, English reading proficiency*, as well as *interaction between the first three effects*) and by L2 (fixed effects of *format, structure, grammaticality, English reading proficiency*, and *interaction between structure and grammaticality*) are given in Table 7 and Table 8.

Table 7

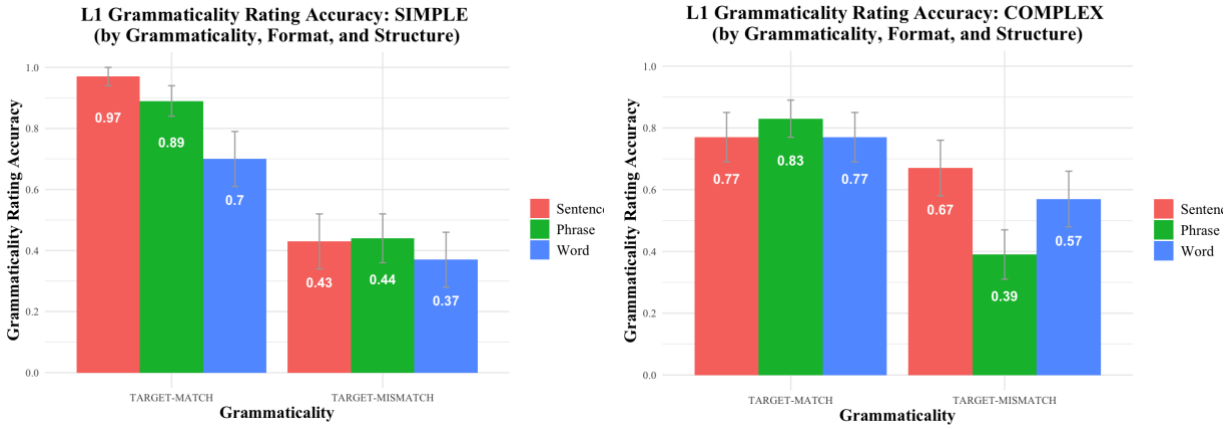
Generalized Linear Mixed-Effects Models of Grammaticality Rating Accuracy by English L1 Speakers. Effects of Format, Structure, Grammaticality, and English Reading Proficiency

Fixed effects	**Estimate**	**S.E**	**z value**	**p value**	**sig**
(Intercept)	0.752	0.346	2.174	0.030	*
Word vs. Sentence	-0.448	0.686	-0.653	0.514	
Phrase vs. Sentence	-0.795	0.630	-1.260	0.208	
Structure	0.267	0.563	0.474	0.635	
Grammaticality	2.553	0.761	3.357	0.001	***
English reading proficiency	0.302	0.147	2.050	0.040	*
Word vs. Sentence x Structure	1.146	1.131	1.013	0.311	
Phrase vs. Sentence x Structure	-0.292	1.137	-0.256	0.798	
Word vs. Sentence x Grammaticality	-0.909	1.704	-0.534	0.594	
Phrase vs. Sentence x Grammaticality	0.661	1.670	0.396	0.692	
Structure x Grammaticality	-1.470	1.090	-1.350	0.177	
Word vs. Sentence x Structure x Grammaticality	3.778	2.165	1.745	0.081	.
Phrase vs. Sentence x Structure x Grammaticality	4.626	2.185	2.117	0.034	*

*RatingAcc ~ Format * Structure * Grammaticality + English reading proficiency + (1 + Structure * Grammaticality | Participant) + (1 + English reading prof. * Format * Structure * Grammaticality | Item)*

Figure 5

Grammaticality Rating Accuracy by English L1 Speakers



Results from the generalized linear mixed models of grammaticality rating accuracy by L1 and L2 participants (Table 7 and Table 8) showed an effect of grammaticality. Since participants were generally biased towards judging sentences as *grammatical*, rating accuracy for the target-match sentences was generally higher than the target-mismatch sentences. Participants with higher English reading proficiency were found to have higher rating accuracy than those less proficient, which was demonstrated in the overall effect of English reading proficiency.

For the native speakers, there was a three-way interaction between format, structure, and grammaticality. L1 participants were better able to correctly identify the mismatch between the anaphor and the reflexive in the experimental items when the sentences were presented in the *whole-sentence* format. However, the *phrase-segment* format was found to be more beneficial to L1 participants than the *sentence* format in the grammaticality judgment of the target-match constructions (see Figure 5).

Table 8

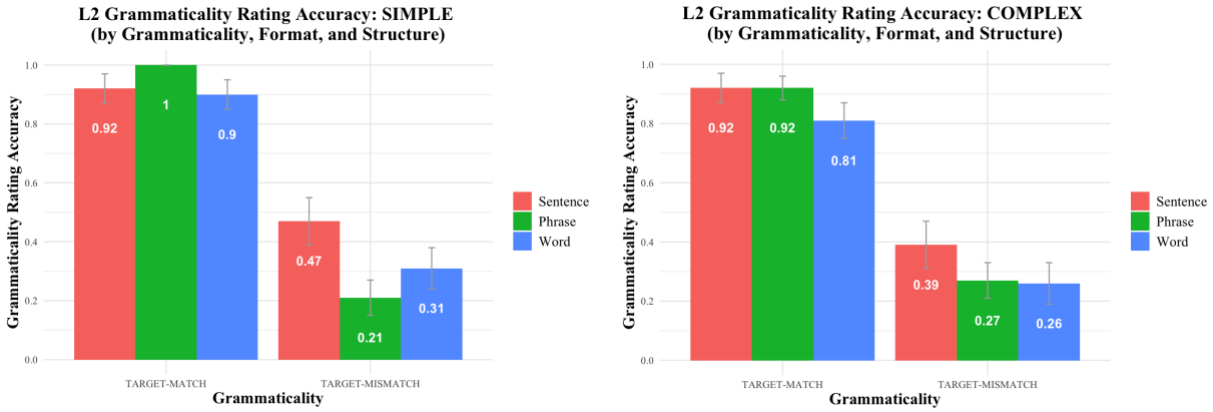
Generalized Linear Mixed-Effects Models of Grammaticality Rating Accuracy by English L2 Speakers. Effects of Format, Structure, Grammaticality, and English Reading Proficiency

Fixed effects	Estimate	S.E	z value	p value	sig
(Intercept)	0.977	0.165	5.905	0.000	***
Word vs. Sentence	-0.724	0.327	-2.210	0.027	*
Phrase vs. Sentence	-0.637	0.325	-1.960	0.050	*
Structure	-0.499	0.292	-1.707	0.088	.
Grammaticality	3.360	0.301	11.150	<0.001	***
English reading proficiency	0.100	0.052	1.909	0.056	.
Structure x Grammaticality	-0.795	0.565	-1.408	0.159	

*RatingAcc ~ Format * Structure * Grammaticality + English reading proficiency + (1 + Structure * Grammaticality | Participant) + (1 + English reading prof. * Format * Structure * Grammaticality | Item)*

Figure 6

Grammaticality Rating Accuracy by English L2 Speakers



For the L2 group, there was an effect of format (see Table 8). Compared with the *word* and *phrase* conditions, L2 participants were most facilitated by the *sentence* condition in the grammaticality judgment task. Findings from the model additionally revealed a marginal effect

of structure among L2 participants, such that L2 participants had marginally higher rating accuracy for simple than for complex constructions.

The overall effect of grammaticality was also demonstrated in the results of the generalized linear mixed models of grammaticality rating accuracy by each of the presentation formats. As participants biased towards *grammatical* sentences (i.e., negative values of c signified participants' tendency to accept the experimental stimuli as *grammatical*), the ratings of grammatical sentences were generally more accurate than the ratings of ungrammatical constructions regardless of the text presentation formats.

Table 9

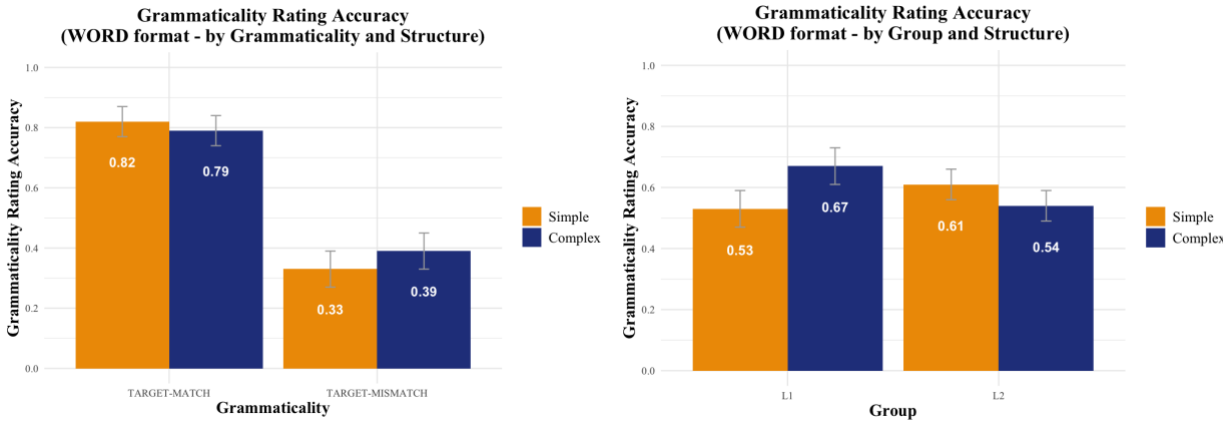
Generalized Linear Mixed-Effects Model of Grammaticality Rating Accuracy in the WORD Presentation Format. Effects of Language Group, Structure, Grammaticality and English Reading Proficiency

Fixed effects	**Estimate**	**S.E**	**z value**	**p value**	**sig**
(Intercept)	0.655	0.259	2.530	0.011	*
Group	0.080	0.474	0.168	0.867	
Structure	0.053	0.398	0.134	0.893	
Grammaticality	2.577	0.596	4.320	0.000	***
English reading proficiency	0.049	0.112	0.440	0.660	
Group x Structure	-1.445	0.707	-2.045	0.041	*
Group x Grammaticality	2.034	1.061	1.917	0.055	.
Structure x Grammaticality	-0.590	0.948	-0.622	0.534	
Group x Structure x Grammaticality	0.046	1.604	0.029	0.977	

*RatingAcc ~ Group * Structure * Grammaticality + English reading proficiency + (1 + Structure * Grammaticality | Participant) + (1 + Group * Structure * Grammaticality * English reading prof. | Item)*

Figure 7

Grammaticality Rating Accuracy in the WORD Presentation Format



Results of the generalized linear mixed model of grammaticality rating accuracy in the *word* format displayed an interaction between group and structure (Table 9). L1 participants had higher grammaticality rating accuracy than L2 for complex constructions, while L2 performed slightly better than L1 for the simple constructions (see Figure 7).

Table 10

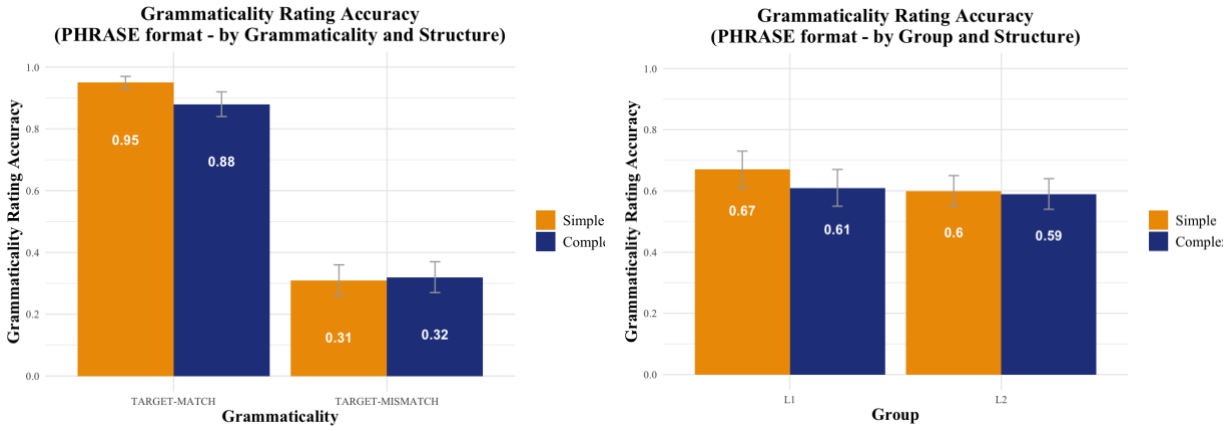
Generalized Linear Mixed-Effects Model of Grammaticality Rating Accuracy in the PHRASE Presentation Format. Effects of Language Group, Structure, Grammaticality and English Reading Proficiency

Fixed effects	Estimate	S.E	z value	p value	sig
(Intercept)	2.335	0.767	3.044	0.002	**
Group	-0.374	0.520	-0.720	0.472	
Structure	-1.813	1.498	-1.211	0.226	
Grammaticality	7.326	1.632	4.490	0.000	***
English reading proficiency	-0.022	0.166	-0.135	0.893	
Structure x Grammaticality	-4.731	2.944	-1.607	0.108	
Structure x English reading proficiency	-0.387	0.339	-1.142	0.253	
Grammaticality x English reading proficiency	-0.438	0.394	-1.111	0.266	
Structure x Grammaticality x English reading proficiency	-0.025	0.640	-0.039	0.969	

*RatingAcc ~ Group * Structure * Grammaticality + English reading proficiency + (1 + Structure * Grammaticality | Participant) + (1 + Group * Structure * Grammaticality * English reading prof. | Item)*

Figure 8

Grammaticality Rating Accuracy in the PHRASE Presentation Format



No other effects or interactions concerning grammaticality rating accuracy in the *phrase-segment* format were found significant, except for the overall effect of grammaticality.

Table 11

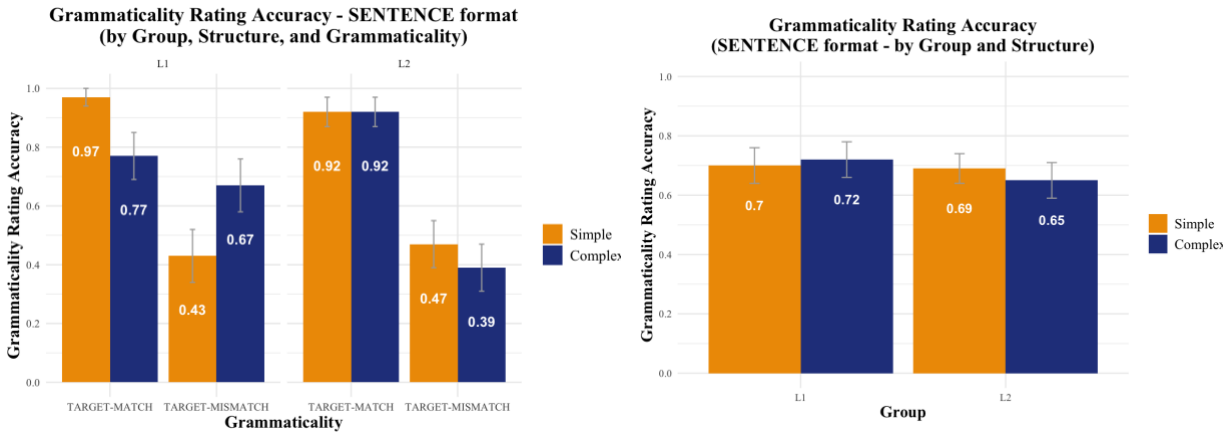
Generalized Linear Mixed-Effects Model of Grammaticality Rating Accuracy in the SENTENCE Presentation Format. Effects of Language Group, Structure, Grammaticality and English Reading Proficiency

Fixed effects	Estimate	S.E	z value	p value	sig
(Intercept)	2.293	1.235	1.856	0.064	.
Group	0.504	1.159	0.435	0.664	
Structure	-1.938	2.235	-0.867	0.386	
Grammaticality	5.057	2.813	1.797	0.072	.
English reading proficiency	0.334	0.195	1.715	0.086	.
Group x Structure	-0.761	1.873	-0.406	0.685	
Group x Grammaticality	1.994	2.812	0.709	0.478	
Structure x Grammaticality	-5.012	4.396	-1.140	0.254	
Group x Structure x Grammaticality	6.947	3.525	1.971	0.049	*

*RatingAcc ~ Group * Structure * Grammaticality + English reading proficiency + (1 + Structure * Grammaticality | Participant) + (1 + Group * Structure * Grammaticality * English reading prof. | Item)*

Figure 9

Grammaticality Rating Accuracy in the SENTENCE Presentation Format



For grammaticality rating accuracy in the *sentence* format, in addition to the marginal effect of grammaticality, there was a three-way interaction between group, structure and grammaticality (see Table 11 and Figure 9). L1 participants were better than L2 in identifying

the mismatch between the anaphor and the reflexive in the complex condition. In contrast, the L2 group had higher rating accuracy for the target-match constructions in the complex condition. However, L2's higher accuracy in the ratings of target-match sentences might be the result of their significant bias towards judging sentences as *grammatical*, which was verified by the results of the t-test comparing the c values (i.e., *bias*) between the two groups ($t = 3.11$, $p = 0.002$).

4.3. Comprehension Accuracy Results

Participants' responses to comprehension questions of grammatical sentences were analyzed using a generalized linear mixed-effects model. The model had fixed effects of *group*, *format*, *structure*, *comprehension tasks* (*general-targeted*, *reflexive-targeted questions*), *English reading proficiency*, as well as *interactions between the first four effects*. The model was fitted with random intercepts and random slopes for both participants and items. Results of the model are given in Table 12 and Figure 10.

Table 12

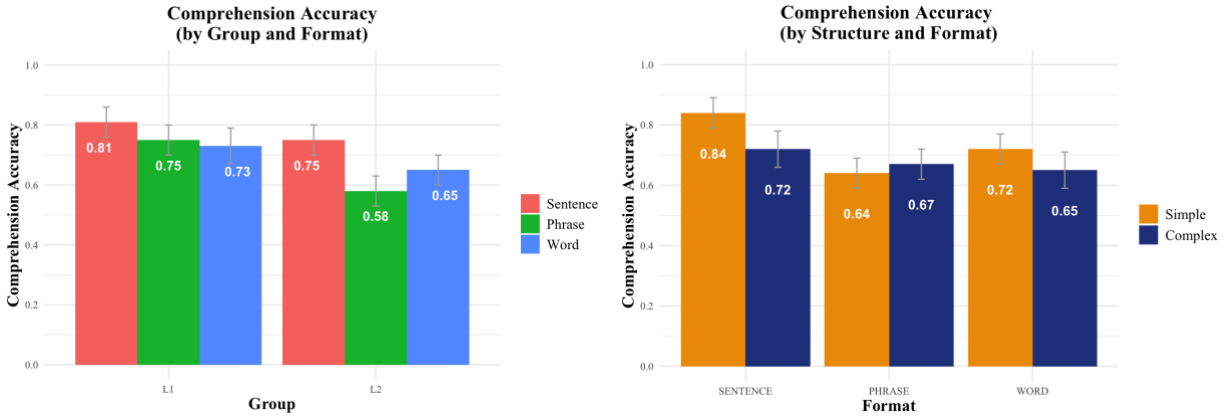
Generalized Linear Mixed-Effects Models of Comprehension Accuracy. Effects of Language Group, Format, Structure, Comprehension Tasks, and English Reading Proficiency

Fixed effects	Estimate	S.E	z-value	p value	sig
(Intercept)	1.149	0.189	6.071	0.000	***
Group	-0.286	0.365	-0.785	0.433	
Word vs. Sentence	-0.689	0.437	-1.579	0.114	
Phrase vs. Sentence	-0.842	0.410	-2.052	0.040	*
Structure	-0.501	0.289	-1.737	0.082	.
Comprehension task	-0.359	0.345	-1.039	0.299	
English reading proficiency	0.085	0.066	1.296	0.195	
Group x Word vs. Sentence	-0.044	0.838	-0.052	0.958	
Group x Phrase vs. Sentence	-0.378	0.788	-0.480	0.631	
Group x Structure	0.104	0.589	0.176	0.860	
Word vs. Sentence x Structure	0.653	0.770	0.848	0.396	
Phrase vs. Sentence x Structure	1.541	0.744	2.072	0.038	*
Group x Comprehension task	-0.125	0.633	-0.198	0.843	
Word vs. Sentence x Comprehension task	-1.919	0.857	-2.240	0.025	*
Phrase vs. Sentence x Comprehension task	-0.953	0.793	-1.202	0.229	
Structure x Comprehension task	-0.370	0.574	-0.644	0.519	
Group x Word vs. Sentence x Structure	0.378	1.695	0.223	0.824	
Group x Phrase vs. Sentence x Structure	-0.969	1.611	-0.602	0.548	
Group x Word vs. Sentence x Comprehension task	1.047	1.652	0.634	0.526	
Group x Phrase vs. Sentence x Comprehension task	-0.676	1.576	-0.429	0.668	
Group x Structure x Comprehension task	1.556	1.185	1.313	0.189	
Word vs. Sentence x Structure x Comprehension task	-2.180	1.536	-1.419	0.156	
Phrase vs. Sentence x Structure x Comprehension task	-0.756	1.474	-0.513	0.608	
Group x Word vs. Sentence x Structure x Comprehension task	-2.170	3.211	-0.676	0.499	
Group x Phrase vs. Sentence x Structure x Comprehension task	-7.898	3.071	-2.572	0.010	*

*Comprehension accuracy ~ Group * Format * Structure * Comprehension Task + English reading proficiency + (1 + Structure + Comprehension Task | Participant) + (1 + Group * Structure * Format | Item)*

Figure 10

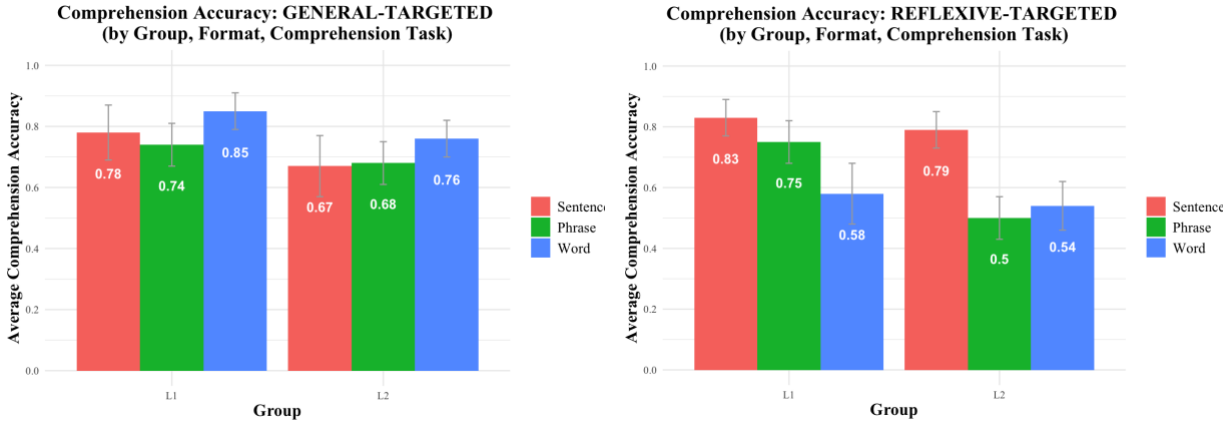
Comprehension Accuracy by Language Group and Format; Structure and Format



On average, L1 participants had higher comprehension accuracy than L2 (*Mean of L1 comprehension accuracy = 0.73, SD = 0.44; Mean of L2 comprehension accuracy = 0.69, SD = 0.46*). There was an overall effect of format, such that participants were most facilitated by the *sentence* format in the comprehension task (Figure 10). For L1 participants, compared to the *phrase* format, the *word* format did not exhibit overt disruption in performance, while for L2, the most disruptive format was the *phrase.* Results in Table 12 and Figure 10 also showed a marginal effect of structure, and an interaction between format and structure, indicating that overall participants had higher comprehension accuracy for simple constructions than for complex constructions. However, comprehension accuracy differed across the three reading formats: accuracy was higher for complex items in the *phrase* condition, while task performance was better across simple constructions in both the *word* and the *sentence* formats.

Figure 11

Comprehension Accuracy by Language Group, Format, and Comprehension Task



The interaction between format and comprehension task demonstrated that the *sentence* format was most beneficial to both groups, manifested in participants' highest accuracy in responses to reflexive-targeted comprehension questions. For L1 participants, the *word* format was most disruptive (though not significantly), while for L2, the least facilitative was the *phrase* format. On average, general-targeted questions were answered more accurately than reflexive-targeted questions. L2 participants had lower comprehension accuracy for reflexive-targeted questions than L1, and particularly so for complex constructions in the *phrase* format, as reflected in the four-way interaction between group, format, structure and comprehension task (Table 12 and Figure 11).

4.4. Reaction Times (RTs) Results

Experimental items of which comprehension questions were answered correctly were selected for analysis of reaction times. Response times were originally recorded in seconds. After being converted into milliseconds and then log-transformed, the data was analyzed using linear mixed-effects models. Proportion of cut-off data was reported in section 4.1.

The following models were fit to the RTs data: (i) linear mixed-effects models of RTs of each segment in the *phrase* format; (ii) linear mixed-effects model of RTs of each sentence in the *whole-sentence* format (with fixed effects of *group*, *structure*, *grammaticality*, *English reading proficiency*, and *all interactions between fixed effects*); (iii) linear mixed-effects model of grammaticality rating RTs (with fixed effects of *group*, *format*, *structure*, *grammaticality*, *English reading proficiency*, and *interactions between the first four effects*); and (iv) linear mixed-effects model of comprehension task RTs (with fixed effects of *group*, *structure*, *grammaticality*, *English reading proficiency*, and *interactions between the first three effects*). All models had random slopes and random intercepts for items and participants.

4.4.1. RTs of Segment 1 in the PHRASE Format

For the *phrase-segment* format, each experimental stimulus was divided into four segments conforming to the syntactic parsing of the sentence.

		Segment 1	**Segment 2**	**Segment 3**	**Segment 4**
Grammatical (Target-match)	**Simple**	The actress	that Mary interviewed at the awards ceremony	about two years ago	described herself as an extreme workaholic.
	Complex	The actress	that Mary interviewed at the awards ceremony	held outside the theater	described herself as an extreme workaholic.
Ungrammatical (Target-mismatch)	**Simple**	The actress	that John interviewed at the awards ceremony	about two years ago	described himself as an extreme workaholic.
	Complex	The actress	that John interviewed at the awards ceremony	held outside the theater	described himself as an extreme workaholic.

Table 13

Linear Mixed-Effects Model of Segment 1 RTs. Effects of Language Group, Structure, Grammaticality, and English Reading Proficiency

Fixed effects	Estimate	S.E	df	t value	p value	sig
(Intercept)	7.105	0.135	24.033	52.692	<0.001	***
Group	0.012	0.269	23.887	0.046	0.964	
Structure	0.012	0.080	34.603	0.152	0.880	
Grammaticality	0.095	0.081	32.843	1.174	0.249	
English reading proficiency	-0.060	0.049	23.900	-1.229	0.231	
Group x Structure	-0.088	0.160	35.157	-0.553	0.584	
Group x Grammaticality	-0.118	0.161	33.053	-0.732	0.470	
Structure x Grammaticality	0.131	0.200	24.401	0.659	0.516	
Group x English reading proficiency	-0.034	0.098	23.972	-0.346	0.733	
Structure x English reading proficiency	-0.003	0.029	35.101	-0.110	0.913	
Grammaticality x English reading proficiency	-0.026	0.030	33.794	-0.880	0.385	
Group x Structure x Grammaticality	0.118	0.397	23.590	0.297	0.769	
Group x Structure x English reading proficiency	-0.017	0.058	34.993	-0.285	0.777	
Group x Grammaticality x English reading proficiency	0.050	0.059	32.648	0.843	0.405	
Structure x Grammaticality x English reading proficiency	0.011	0.073	23.740	0.154	0.879	
Group x Structure x Grammaticality x English reading proficiency	0.106	0.147	22.982	0.722	0.478	

*LogRTSeg1 ~ Group * Structure * Grammaticality * English reading proficiency + (1 + Structure * Grammaticality | Participant) + (1 + Group : Structure : Grammaticality : English reading proficiency | Item)*

Results of the model showed that there were no significant effects or interactions for the RTs of segment 1 in the *phrase* format.

4.4.2. RTs of Segment 2 in the PHRASE Format

		Segment 1	Segment 2	Segment 3	Segment 4
Grammatical (Target-match)	**Simple**	The actress	that Mary interviewed at the awards ceremony	about two years ago	described herself as an extreme workaholic.
	Complex	The actress	that Mary interviewed at the awards ceremony	held outside the theater	described herself as an extreme workaholic.
Ungrammatical (Target-mismatch)	**Simple**	The actress	that John interviewed at the awards ceremony	about two years ago	described himself as an extreme workaholic.
	Complex	The actress	that John interviewed at the awards ceremony	held outside the theater	described himself as an extreme workaholic.

Table 14

Linear Mixed-Effects Model of Segment 2 RTs. Effects of Language Group, Structure, Grammaticality, and English Reading Proficiency

Fixed effects	**Estimate**	**S.E**	**df**	**t value**	**p value**	**sig**
(Intercept)	7.901	0.061	41.534	130.582	<0.001	***
Group	0.049	0.113	23.098	0.432	0.670	
Structure	0.131	0.046	88.367	2.819	0.006	**
Grammaticality	0.045	0.038	38.004	1.167	0.251	
English reading proficiency	-0.044	0.021	22.219	-2.065	0.051	.
Group x Structure	-0.013	0.063	40.706	-0.210	0.835	
Group x Grammaticality	0.029	0.067	30.982	0.436	0.666	
Structure x Grammaticality	-0.070	0.095	64.445	-0.736	0.465	
Group x Structure x Grammaticality	-0.043	0.135	25.764	-0.319	0.753	

*LogRTSeg2 ~ Group * Structure * Grammaticality + English reading proficiency + (1 + Structure * Grammaticality | Participant) + (1 + Group : Structure : Grammaticality : English reading proficiency | Item)*

As shown in Table 14, there was an effect of structure in the RTs of segment 2, such that participants spent more time reading segment 2 of complex sentences than that of simple sentences. There was also a marginal effect of English reading proficiency, which suggested that

participants with higher English reading proficiency were associated with faster reading times. No other findings were significant.

4.4.3. RTs of Segment 3 in the PHRASE Format

		Segment 1	Segment 2	Segment 3	Segment 4
Grammatical (Target-match)	**Simple**	The actress	that Mary interviewed at the awards ceremony	about two years ago	described herself as an extreme workaholic.
	Complex	The actress	that Mary interviewed at the awards ceremony	held outside the theater	described herself as an extreme workaholic.
Ungrammatical (Target-mismatch)	**Simple**	The actress	that John interviewed at the awards ceremony	about two years ago	described himself as an extreme workaholic.
	Complex	The actress	that John interviewed at the awards ceremony	held outside the theater	described himself as an extreme workaholic.

Table 15

Linear Mixed-Effects Model of Segment 3 RTs. Effects of Language Group, Structure, Grammaticality, and English Reading Proficiency

Fixed effects	Estimate	S.E	df	t value	p value	sig
(Intercept)	7.658	0.058	40.497	132.373	<0.001	***
Group	0.062	0.107	25.865	0.577	0.569	
Structure	0.185	0.054	195.778	3.419	0.001	***
Grammaticality	-0.070	0.047	104.888	-1.488	0.140	
English reading proficiency	-0.048	0.020	26.352	-2.367	0.026	*
Group x Structure	0.277	0.080	249.925	3.462	0.001	***
Group x Grammaticality	-0.113	0.084	79.449	-1.346	0.182	
Structure x Grammaticality	0.116	0.111	159.764	1.048	0.296	
Group x Structure x Grammaticality	0.030	0.164	127.502	0.180	0.857	

*LogRTSeg3 ~ Group * Structure * Grammaticality + English reading proficiency + (1 + Structure * Grammaticality | Participant) + (1 + Group : Structure : Grammaticality : English reading proficiency | Item)*

The syntactic manipulation of the experimental items was reflected in Segment 3 (i.e., the adverbial phrase in the simple configuration was replaced by the past participle clause in the complex configuration). Increased syntactic complexity correlated with increased reading times,

which was manifested in a significant effect of structure. Compared to high proficient participants, those with lower proficiency needed more time to read Segment 3. An interaction between group and structure further indicated that Segment 3 of complex sentences were read at lower speed, particularly so by L2 participants.

4.4.4. RTs of Segment 4 in the PHRASE Format

		Segment 1	Segment 2	Segment 3	Segment 4
Grammatical (Target-match)	**Simple**	The actress	that Mary interviewed at the awards ceremony	about two years ago	described herself as an extreme workaholic.
	Complex	The actress	that Mary interviewed at the awards ceremony	held outside the theater	described herself as an extreme workaholic.
Ungrammatical (Target-mismatch)	**Simple**	The actress	that John interviewed at the awards ceremony	about two years ago	described himself as an extreme workaholic.
	Complex	The actress	that John interviewed at the awards ceremony	held outside the theater	described himself as an extreme workaholic.

Table 16

Linear Mixed-Effects Model of Segment 4 RTs. Effects of Language Group, Structure, Grammaticality, and English Reading Proficiency

Fixed effects	Estimate	S.E	df	t value	p value	sig
(Intercept)	8.014	0.049	29.949	162.067	<0.001	***
Group	0.074	0.103	24.930	0.720	0.478	
Structure	0.152	0.046	39.958	3.293	0.002	**
Grammaticality	-0.140	0.039	27.999	-3.558	0.001	**
English reading proficiency	-0.016	0.020	24.741	-0.821	0.420	
Group x Structure	0.035	0.080	23.278	0.433	0.669	
Group x Grammaticality	0.170	0.074	23.382	2.304	0.030	*
Structure x Grammaticality	0.200	0.089	44.406	2.256	0.029	*
Group x Structure x Grammaticality	0.028	0.152	26.869	0.185	0.855	

*LogRTSeg4 ~ Group * Structure * Grammaticality + English reading proficiency + (1 + Structure * Grammaticality | Participant) + (1 + Group : Structure : Grammaticality : English reading proficiency | Item)*

Segment 4 contained the critical region (i.e., reflexive). Results of the mixed model revealed that complexity interfered with the reading times of segment 4. In other words, compared with simple sentences, participants spent more time reading segment 4 of complex sentences. The mismatch between the reflexive and the anaphor also led to increased reading times of the segment, which was reflected in an overall effect of grammaticality. The interaction between group and grammaticality signified the difference between the two groups in the reading times of segment 4 in constructions that varied in grammaticality. Specifically, L2 participants needed more time to read segment 4, particularly segment 4 in grammatical sentences. Grammaticality also interacted with structure, suggesting that for grammatical sentences, segment 4 of complex constructions took participants longer time to read than that of simple constructions.

4.4.5. RTs of Sentences in the SENTENCE Format

Table 17

Linear Mixed-Effects Model of Sentence RTs. Effects of Language Group, Structure, Grammaticality, and English Reading Proficiency

Fixed effects	Estimate	S.E	df	t value	p value	sig
(Intercept)	9.327	0.063	18.941	148.314	<0.001	***
Group	0.152	0.151	19.119	1.005	0.327	
Structure	0.076	0.034	18.196	2.204	0.041	*
Grammaticality	-0.023	0.033	19.188	-0.702	0.491	
English reading proficiency	-0.056	0.026	18.957	-2.175	0.043	*
Group x Structure	-0.085	0.068	17.335	-1.251	0.228	
Group x Grammaticality	0.055	0.066	19.762	0.839	0.412	
Structure x Grammaticality	-0.026	0.060	20.023	-0.426	0.675	
Group x Structure x Grammaticality	0.108	0.119	19.200	0.913	0.372	

*LogRTSentence ~ Group * Structure * Grammaticality + English reading proficiency + (1 + Structure * Grammaticality | Participant) + (1 + Group : Structure : Grammaticality : English reading proficiency | Item)*

Results of the mixed model of *sentence* RTs displayed an effect of structure. Increased syntactic complexity resulted in an increase in total RTs of the sentence. Participants with higher English reading proficiency read sentences faster than those with lower proficiency. No other findings were significant.

4.4.6. RTs of Grammaticality Rating Task

Participants' response times in the grammaticality rating task were reported in Table 18. Findings from the mixed model demonstrated that there was an effect of format in the RTs of grammaticality rating task. In contrast to the *sentence* format, the *word* and the *phrase* formats were associated with increased RTs in grammaticality rating. An effect of structure suggested that in general, participants spent less time in the rating task of complex sentences; however, an interaction between structure and grammaticality further indicated that for grammatical constructions, participants needed more time to rate complex sentences than simple sentences.

Table 18

Linear Mixed-Effects Model of Grammaticality Rating RTs. Effects of Language Group, Format, Structure, Grammaticality, and English Reading Proficiency

Fixed effects	Estimate	S.E	df	t value	p value	sig
(Intercept)	7.385	0.038	70.654	195.177	<0.001	***
Group	-0.082	0.083	66.046	-0.982	0.330	
Word vs Sentence	0.247	0.091	63.802	2.701	0.009	**
Phrase vs Sentence	0.250	0.088	64.204	2.848	0.006	**
Structure	-0.079	0.037	101.706	-2.163	0.033	*
Grammaticality	-0.030	0.050	65.908	-0.594	0.555	
English reading proficiency	-0.002	0.016	64.195	-0.108	0.914	
Group x Word vs Sentence	-0.183	0.184	65.900	-0.996	0.323	
Group x Phrase vs Sentence	-0.175	0.174	64.721	-1.002	0.320	
Group x Structure	-0.006	0.079	68.278	-0.079	0.937	
Word vs Sentence x Structure	-0.062	0.096	112.037	-0.644	0.521	
Phrase vs Sentence x Structure	-0.146	0.093	112.102	-1.570	0.119	
Group x Grammaticality	-0.010	0.103	70.321	-0.101	0.920	
Word vs Sentence x Grammaticality	0.038	0.133	70.057	0.285	0.777	
Phrase vs Sentence x Grammaticality	0.092	0.127	69.851	0.727	0.470	
Structure x Grammaticality	0.176	0.073	116.517	2.422	0.017	*
Group x Word vs Sentence x Structure	-0.113	0.188	103.526	-0.603	0.548	
Group x Phrase vs Sentence x Structure	-0.059	0.185	94.474	-0.318	0.751	
Group x Word vs Sentence x Grammaticality	0.123	0.259	69.732	0.476	0.636	
Group x Phrase vs Sentence x Grammaticality	0.089	0.247	68.247	0.360	0.720	
Group x Structure x Grammaticality	-0.157	0.147	119.994	-1.073	0.286	
Word vs Sentence x Structure x Grammaticality	0.063	0.191	127.066	0.327	0.744	
Phrase vs Sentence x Structure x Grammaticality	-0.180	0.185	124.567	-0.973	0.333	
Group x Word vs Sentence x Structure x Grammaticality	-0.264	0.422	40.227	-0.625	0.535	
Group x Phrase vs Sentence x Structure x Grammaticality	-0.133	0.382	54.887	-0.347	0.730	

*LogRating ~ Group * Format * Structure * Grammaticality + English reading proficiency + (1 + Structure * Grammaticality | Participant) + (1 + Group : Format : Structure : Grammaticality | Item)*

4.4.7. RTs of Comprehension Task

Table 19

Linear Mixed-Effects Model of Comprehension Task RTs. Effects of Language Group, Format, Structure, Comprehension Task, and English Reading Proficiency

Fixed effects	Estimate	S.E	df	t value	p value	sig
(Intercept)	8.046	0.032	62.547	250.422	<0.001	***
Group	0.054	0.063	62.546	0.855	0.396	
Word vs Sentence	-0.062	0.074	71.191	-0.834	0.407	
Phrase vs Sentence	0.125	0.070	68.055	1.782	0.079	.
Structure	0.001	0.033	112.117	0.017	0.987	
Comprehension task	-0.142	0.061	64.991	-2.317	0.024	*
English reading proficiency	-0.048	0.012	63.799	-3.923	<0.001	***
Group x Word vs Sentence	-0.208	0.140	61.228	-1.490	0.141	
Group x Phrase vs Sentence	-0.250	0.132	60.996	-1.903	0.062	.
Group x Structure	-0.031	0.069	82.721	-0.459	0.648	
Word vs Sentence x Structure	-0.061	0.102	105.820	-0.602	0.549	
Phrase vs Sentence x Structure	-0.015	0.100	90.365	-0.149	0.882	
Group x Comprehension task	-0.022	0.103	61.646	-0.215	0.830	
Word vs Sentence x Comprehension task	0.129	0.145	77.622	0.894	0.374	
Phrase vs Sentence x Comprehension task	0.188	0.136	71.946	1.390	0.169	
Structure x Comprehension task	-0.001	0.067	119.227	-0.012	0.990	
Group x Word vs Sentence x Structure	0.229	0.184	40.320	1.245	0.220	
Group x Phrase vs Sentence x Structure	0.029	0.168	72.687	0.174	0.862	
Group x Word vs Sentence x Comprehension task	0.154	0.269	66.150	0.574	0.568	
Group x Phrase vs Sentence x Comprehension task	-0.463	0.251	62.628	-1.845	0.070	.
Group x Structure x Comprehension task	-0.346	0.137	86.984	-2.529	0.013	*
Word vs Sentence x Structure x Comprehension task	0.639	0.203	101.633	3.148	0.002	**
Phrase vs Sentence x Structure x Comprehension task	0.329	0.199	104.865	1.651	0.102	
Group x Word vs Sentence x Structure x Comprehension task	-0.471	0.367	39.094	-1.284	0.207	
Group x Phrase vs Sentence x Structure x Comprehension task	-0.456	0.336	82.983	-1.360	0.177	

*LogRTComp ~ Group * Structure * Grammaticality + English reading proficiency + (1 + Structure * Grammaticality | Participant) + (1 + Group : Structure : Grammaticality : English reading proficiency | Item)*

An effect of comprehension task indicated that participants differed in the amount of time needed to answer different types of questions in the comprehension task. Questions that targeted the relation between the anaphor and the reflexive (i.e., reflexive-targeted questions) took participants longer time to answer than questions that asked about other information of the sentence (i.e., general-targeted questions). There was a significant effect of English reading proficiency, such that high proficient participants did not need to spend as much time as those less proficient in responding to comprehension probes.

The three-way interaction between group, structure, and comprehension task suggested L1-L2 difference in the RTs of reflexive-targeted questions: compared to L1, L2 participants spent less time responding to reflexive-targeted questions for complex sentences. Finally, an interaction between format, structure, and comprehension task further indicated the negative interference of the *word* format in participants' response times of the comprehension questions. It took participants longer time to respond to reflexive-targeted questions of complex sentences, and especially so when the sentences were presented in the *word* format. No other effects or interactions were significant.

4.5. Evidence for the Application of a Good-enough Processing Strategy

To investigate whether participants showed evidence of the application of a good-enough processing strategy, I analyzed the experimental items where comprehension questions targeted general information. The exclusion of reflexive-targeted questions in the analysis of good-enough processing eliminated the inclusion of items of which grammaticality was accurately rated, but contradictorily, reflexive-targeted questions were answered inaccurately. Correlation tests between general comprehension accuracy and grammaticality rating accuracy were performed to examine whether there were any trade-off effects between participants'

general comprehension and their ability to detect grammatical anomalies. Manipulation of syntactic complexity was taken into consideration to get insight into how increased cognitive load, resulting from increased syntactic complexity, might interfere with participants' accuracy in syntactic analysis.

Results of the correlation test between general comprehension accuracy and grammaticality rating accuracy for the simple constructions (*r = 0.12; n = 74; p = 0.36*) did not reveal any significant correlation between general comprehension and grammaticality judgment. However, a significantly negative correlation (*r = -0.27; n = 74; p = 0.02*) between general comprehension accuracy and grammaticality rating accuracy was found among sentences in the complex configuration.

4.6. Summary of Findings

4.6.1. Grammaticality Judgment

Participants, in general, biased towards accepting the experimental sentences as *grammatical*. Compared with the L1 group, the nonnative participants tended to give higher ratings to the experimental items, which resulted in L2's higher grammaticality rating accuracy for the target-match sentences. However, on average, L1 participants showed higher sensitivity for grammatical and ungrammatical sentences than L2. In other words, L1 participants were better able to differentiate between target-match and target-mismatch sentences.

Overall, the *sentence* format was most facilitative for both groups in the grammaticality judgment task. Syntactic complexity also affected rating accuracy, such that sentences, particularly grammatical sentences, in the complex condition were rated less accurately than sentences in the simple condition. Higher English reading proficiency was associated with higher grammaticality rating accuracy.

4.6.2. Comprehension Accuracy

Both groups were most facilitated by the *sentence* format in the comprehension task. However, the two groups differed in task performance across the other two text presentation formats. While the *word* format was most disruptive for the L1 group, for the L2, it was the *phrase*.

On average, L1 participants had higher comprehension accuracy than L2; simple sentences had higher comprehension accuracy than complex sentences; and general-targeted questions were answered more accurately than reflexive-targeted questions.

4.6.3. Reaction Times

Reading times of the four segments in the *phrase-segment* format, and reading times of the sentences in the *sentence* format were slower for complex than for simple constructions. There were increased reading times at segment 3 (i.e., manipulation of syntactic complexity) and segment 4 (i.e., the critical region containing the reflexive). Participants with higher English reading proficiency generally had faster reading times than participants with lower proficiency.

Experimental sentences presented in the *word* and the *phrase* format had longer reaction times in the grammaticality rating task compared to the sentences presented in the *sentence* paradigm. Participants needed more time to rate complex target-match sentences than simple target-match sentences. However, generally, complex sentences took participants less time to rate than simple sentences.

Reflexive-targeted questions were answered more quickly than general-targeted questions. On the whole, participants who scored better in the reading proficiency test answered comprehension questions faster than those who scored lower. L2 participants responded to reflexive-targeted questions of the complex items faster than L1 participants. Compared to the

sentence condition, reflexive-targeted questions for complex constructions in the *word* condition took participants longer time to answer.

4.6.4. Good-enough Processing

There was evidence of good-enough processing when syntactic complexity increased. Tradeoff effects between general comprehension and ability to detect grammatical errors were observed among complex constructions, but not among simple constructions. Alternatively stated, general comprehension accuracy remained high despite manipulation of syntactic complexity, while grammaticality rating accuracy was negatively affected when there was an increase in cognitive load.

Chapter 5. Discussion

This chapter discusses the results of the study in light of the three areas of inquiry that constitute the suggested prosody-memory integrated model for sentence processing (Pratt & Fernández, 2016): (i) cue-based memory retrieval (Lewis & Vasishth, 2005; Engelmann et al., 2019), (ii) implicit prosody (Bader, 1998; Fodor, 1998, 2002), (iii) Good-enough Processing hypothesis (Ferreira, et al., 2002; Ferreira, 2003; Ferreira & Patson, 2007) and Online Cognitive Equilibrium hypothesis (Karimi & Ferreira, 2016).

5.1. Cue-based Memory Retrieval in Anaphora Resolution

Accumulating body of research has pointed out that interference effects also occurred in reflexive-antecedent licensing (e.g., Patil et al., 2016; Parker & Phillips, 2017), and that memory access mechanism was applied in both subject-verb agreement and reflexive-antecedent dependencies. During processing of anaphoric reflexives, parsers are expected to make use of both structural and non-structural cues for retrieval of the target antecedent. Spreading of cue activation as a result of interference effects was found to lead to reduction in the association strength between the cues and the target, which consequently gives rise to a delay in the retrieval of the target antecedent (Patil et al., 2016). Though it was suggested that the weighting of structural and nonstructural cues might determine the extent to which attraction effects interfere with the retrieval process, questions remained open as to which specific cue weightings that comprehenders prioritize during retrieval, and if a single retrieval probe is created by multiple cues via a linear or non-linear method of cue combination (Parker, 2019).

Findings of most studies are in line with the mechanism of a linear memory access model such as the LV05 (Lewis & Vasishth, 2005). However, in order to better understand the cognitive processes involved in sentence comprehension, the development of a model that

adopts either a nonlinear method (Parker, 2019) or a mixed (i.e., combination of linear and nonlinear) method still calls for further research. Recently, cue combinatorics has been revisited to get further insights in the cognitive processes involved in sentence processing and comprehension (e.g., Engelmann, et al., 2019; Parker, 2019). Findings from Parker's (2019) study are in favor of a cognitive domain that incorporates both linear and nonlinear cue combination methods. Interference effects in reflexive licensing which could not be detected via a linear cue combination method could now be observed via a nonlinear one (Parker, 2019). The necessity of retrieval models that engage both linear and nonlinear cue combination rule was claimed to stem from the differences between subject-verb agreement and anaphor-reflexive relation in resolution of dependencies. However, up to date, little is known with respect to how cue weighting affects cue combinatorics in retrieval of cues during sentence processing. Within the scope of this study, only the evidence from earlier research concerning cue combinatorics will be linked to the discussion of the present data. Issues or domains that have not been thoroughly researched (i.e., cue reliability, cue prediction, relationship between retrieval and response times, etc.) are left for future work.

Though further research is needed, the computational model of language processing – the revised LV05+IP+MAC model (Lewis & Vasishth, 2005; Engelmann et al., 2019) – still provides sound explanations for the interference effects that other preceding models failed to capture or justify due to simplification of model assumptions. Further discussion on how the revised LV05 model (Engelmann et al., 2019) fits with the data of the present study will be provided in section 5.4.1.

5.2. Good-enough Processing and Cognitive Equilibrium in Anaphora Resolution

The Good-enough Processing hypothesis (Ferreira, et al., 2002; Ferreira, 2003; Ferreira & Patson, 2007) and the Online Cognitive Equilibrium hypothesis (Karimi & Ferreira, 2016) suggested that during processing comprehenders have a tendency to favor a simple, heuristic processing strategy over an effortful syntactic analysis despite potential cost of sentence misinterpretation. Only when required by the task, an algorithmic route to processing is deployed. Misretrieval is likely to occur, especially in cases where the cue matches the feature of the inaccessible antecedent but does not match that of the target. On the contrary, comprehension accuracy still remained relatively high (Patil et al., 2016), which signifies the absence of a deep, algorithmic route of processing among comprehenders (when other factors such as working memory capacity, inattentiveness, etc. are held constant). Decision on the application of a processing route largely depends on the requirement of the task at hand and/or the differences among individuals (von der Malsburg & Vasishth, 2013; Traxler, 2007). Unless explicitly mandated by the task, parsers, regardless of their language profiles, tend to settle on a shallower approach which is "good-enough" to uphold general comprehension (Pratt & Fernández, 2016; Swets, Desmet, Clifton, & Ferreira, 2008; Tan & Foltz, 2020).

Participants of the present study were assigned the comprehension task that either required reflexive licensing or recollection of general information. Moreover, besides task difficulty, manipulation of the prosodic contours projected onto the experimental sentences were hypothesized to increase cognitive load among participants, which was expected to prompt participants to lean toward a more resource-preserving route of processing. How a good-enough approach was adopted by participants under high task demand will be discussed in depth in section 5.4.2.

5.3. Implicit Prosody in Anaphora Resolution

The Implicit Prosody hypothesis (IPH) (Bader, 1998; Fodor, 1998, 2002) suggested that the prosodic parsing induced during silent reading might facilitate resolution of syntactic ambiguity. Over the past 20 years, the application of the IPH has been featured in studies on ambiguity resolution including garden path sentences and relative clause attachments, etc. (e.g., Bader, 1998; Fodor, 1998, 2002; Fernández et al., 2003; Hirose, 2003; Hwang & Schafer, 2009; Jun & Bishop, 2015; Swets et al., 2008; Traxler, 2009; Webman-Shafran, 2018; Yao & Scheepers, 2018). Prior to the Implicit Prosody hypothesis, there had already been studies on how pitch contours and durational information tied to syntactic structures affected the metalinguistic analysis of a sentence (e.g., Beach, 1991). Findings from recent studies using sensitive techniques such as eye-tracking and event-related potentials have provided further evidence for the effects of prosody during the earlier stage of parsing (Ferreira & Cokal, 2008). There has also been increasing consensus that prosodic parsing in line with syntactic boundaries eases comprehension by reducing cognitive load (Kreiner, 2005; Pratt & Fernández, 2016). Syntactic cues extracted from corresponding prosodic contours were suggested to guide the processing and/or the reanalysis of a sentence, thus possibly preventing comprehenders from sentence misinterpretation.

Concerning the projection of prosody onto syntactic constructions that contain anaphors, Wolters and Byron (2000) found evidence of individual differences on the application of prosodic marking in the resolution of personal and demonstrative pronouns, and that the effects of prosodic cues in anaphora resolution were not robust (at least in the materials of their study) if other factors such as dialog acts and dialog structure were not taken into consideration.

The sentence processing model that integrated implicit prosody and cue-based memory retrieval suggested by Pratt and Fernández (2016) was among the first of its kind that looked into how clausal edges act as a guiding cue, and how implicit prosody that deviates from clausal edges might be costly for L1 and L2 speakers in the processing of relative clause attachments. Moreover, syntactic complexity is expected to negatively interfere with processing, since processing under duress (i.e., disruptive prosodic parsing, increased syntactic complexity) places a heavy toll on cognitive load, which might result in participants' resort to a good-enough processing strategy.

This study extends Pratt and Fernández's (2016) prosody-memory integrated model to the processing of syntactic constructions that contain anaphoric reflexives. The purpose of the study is to examine if a similar pattern of results would be obtained with respect to the processing of anaphoric reflexives. In other words, if anaphora resolution would be affected by manipulations of syntactic complexity and implicit prosody, and if there would be any similarities and/or differences in L1 and L2 processing of reflexive-antecedent dependencies.

5.4. Discussion of Results

Given the purpose of the study, this section discusses the findings with respect to the application of (i) Cue-based Memory Retrieval mechanism, (ii) Implicit Prosody hypothesis, and (iii) Good-enough Processing and Online Cognitive Equilibrium hypotheses in the processing and resolution of anaphoric reflexives.

5.4.1. Cue-based Memory Retrieval in the Resolution of Anaphoric Reflexives

The experimental items of the current study were adapted from the stimuli used in Patil et al.'s (2016) research. Distractors were assigned the same syntactic role (subject role) as the target antecedent. Manipulation of grammaticality was reflected in the gender match/mismatch

between the target and the reflexive, while the distractors in the two configurations (i.e., match and mismatch) always gender-matched the reflexive (see Table 2 for a sample set of experimental items). Interference effects in the match and mismatch configurations, thus, were expected to differ.

There has been growing consensus over interference effects in reflexive binding in English (e.g., Patil et al., 2016; Parker & Phillips, 2017). Syntactic constraints (i.e., Binding Principle A) were suggested by earlier research to be the primary guiding cue in the retrieval of antecedent, reflected in the parser's negligence of items that were not syntactically legitimate for retrieval (e.g., Dillon et al., 2013; Xiang et al., 2009). Findings from more recent research pointed out otherwise, such that besides structural cues, parsers also make use of nonstructural cues in the search for a grammatically accessible antecedent (e.g., Patil et al., 2016; Parker & Phillips, 2017). Detection of interference effects in reflexive binding was argued to depend on the match in grammatical role between the target and the distractor(s) (Engelmann et al., 2019; Patil et al., 2016; Van Dyke & McElree, 2011).

According to the original LV05 model (Lewis & Vasishth, 2005) and the revised LV05+IP+MAC model (Engelmann et al., 2019), matching features between retrieval candidates are predicted to lead to spreading activation across items (i.e., fan effect) even in the target-mismatch configurations, which in turn results in longer retrieval latency, and sometimes misretrieval of antecedent. As for this study, it is predicted that: (i) interference effects will be more robust in L2 than in L1 participants; (ii) and the effects are expected to be most robust in the complex/target-mismatch condition when syntactic complexity interacts with grammaticality.

Regarding the first prediction, results from the overall sensitivity test showed that on average L1 participants showed higher sensitivity than L2. In other words, the L1 group was better able to distinguish between grammatical and ungrammatical sentences than the L2. In the *word* condition, L1 significantly showed higher grammaticality rating accuracy than L2, and particularly so for complex constructions. Similar results were observed in the *sentence* condition, reflected in L1's higher rating accuracy in complex/target-mismatch configurations. The findings suggest that compared to L1 participants, L2 were more prone to interference effects, which supports the claim made by Pratt and Fernández (2016) that attraction effects were stronger in L2 than in L1[3].

The second prediction accounts for findings from past research that found higher rate of retrieval errors when the accessible antecedent and the reflexive do not match in gender (e.g., Patil et al., 2016; Sturt, 2003). For the second prediction, recency and item prominence were predicted by the ACT-R theory and the updated LV05 model to associate with memory decay, which, as a result, leads to reduction in the activation strength of an item (Jager, Benz, Roeser, Dillon, & Vasishth, 2015). In addition, the extended LV05 model further predicts reactivation boosts in items with high level of prominence (Engelmann et al., 2019). As for recency, when the matching distractor is linearly closer to the reflexive than the mismatching target, the matching distractor will have a recency advantage, which consequently leads to higher probability of misretrieval (Jager et al., 2015). The second prediction expects the strongest interference effects in complex/target-mismatch configurations when there is an interaction between complexity and grammaticality. With respect to the effect of complexity, results from the sensitivity-and-bias-by-item test revealed that participants marginally had higher sensitivity

[3] In Pratt and Fernández's (2016) study, the syntactic configuration was relative clauses.

to grammatical sentences in the simple condition as compared to the grammatical sentences in the complex condition. The findings were further supported by the interaction between structure and grammaticality which was reported in the results of the mixed model of grammaticality rating accuracy (see Table 7). Even though overall, participants could differentiate between target-match and target-mismatch sentences as pointed out by the overall sensitivity test, the sensitivity participants had for the target-mismatch condition in particular was significantly low. In other words, compared to the target-match configuration, the rate of misretrieval was remarkably higher for the target-mismatch configuration. This finding is in line with one of the predictions of the ACT-R theory, and also with findings from past research which found evidence of mismatch effect in retrieval times as well as retrieval errors (i.e., longer total reading time at the reflexive, and higher rates of misretrieval when the target and the reflexive did not match in gender) (Patil et al., 2016; Sturt, 2003).

For L1 participants, the prosodic contours that align with the clausal edges of the sentence, as depicted in the *phrase-segment* format, facilitated the processing of complex/target-match sentences, while for L2, the *phrase* format was most beneficial in the processing of simple/target-match sentences. The results relate to one of the findings from Pratt and Fernández's (2016) research which attributed L1 participants' ability to distinguish simple/grammatical RC constructions to the alignment between a sentence's prosodic parsing and its syntactic clausal edges. The *phrase* format was predicted by Pratt and Fernández to possibly slow down cue decay and provide a stronger activation boost to the accessible target. However, as reported in the Results chapter, other than the aforementioned findings, the current data did not observe any other significant effects of the *phrase* format in the grammaticality rating task. In fact, participants' performance in the *sentence* format was generally better than

their performance in the *phrase* format in terms of both sensitivity and accuracy rate of grammaticality judgment. The findings put forward a suggestion that although prosodic contours in line with syntactic parsing did help participants with processing under certain circumstances, there still exist other scenarios in which the alignment of prosody and syntax was not at all facilitative.

The analysis of the reaction times of segment 4 (i.e., the critical region that contains the reflexive) indicated the effects of structure and grammaticality, such that participants needed more time to read segment 4 of the complex sentences as compared to that of the simple sentences. Participants also spent longer time on the critical segment of the target-mismatch than that of the target-match condition. Similar results were found in some other studies (e.g., Cunnings & Felser, 2013; Cunnings & Sturt, 2014; Parker & Phillips, 2014; Patil et al., 2016; Sturt, 2003) which showed the mismatch effect in retrieval times: longer total reading time at the reflexive, and longer retrieval latency in the target-mismatch than in the target-match condition.

Concerning the analysis of distractor prominence, cue weights, and spreading activation of items for the current data, I applied the cue combinatorics suggested by the revised LV05 model (Engelmann et al., 2019) which takes into account the extended assumptions of item prominence and multi-associative cues. As the revised LV05 predicts that "a cue can have variable discrimination", and that "it can be associated with multiple features to different degrees" (Engelmann et al., 2019, p. 25), spreading activation of the items (i.e., target and distractors) in the experimental stimuli of this study were calculated with respect to the associative strength between the cues and the features of the items. Item prominence will also be discussed in depth since the distractors in both the target-match and target-mismatch conditions

had the grammatical role similar to that of the target. As grammaticality was manipulated across experimental sentences, item prominence, cue weights, and spreading activation of items will be computed for each configuration (i.e., target-match, target-mismatch) in separation.

5.4.1.1. Multi-associative Cues

5.4.1.1.1. Target-mismatch Configuration

The actress$_{+CCOM}^{-MASC}$ that

John$_{-CCOM}^{+MASC}$ interviewed at the awards ceremony about two years ago described

himself$\{_{CCOM}^{MASC}\}$ as an extreme workaholic.[4]

The target item *actress* only has the feature *+CCOM* that matches the cue CCOM of the reflexive *himself*; and the distractor item *John* only has the feature *+MASC* that matches the cue MASC of the reflexive. It was assumed by the LV05 model that the two cues, MASC and CCOM, would be utilized during the retrieval of the reflexive. Since there is no cross-association of features between the target and the distractor in the target-mismatch condition, the spreading activation from the cue CCOM to the target, and from the cue MASC to the distractor will be calculated respectively as follows.

a. Spreading activation of target

The notations below are used in the equations discussed in this section:

- *v*: all active memory items given cue CCOM
- K_i: all features of the item *i*
- $Q_{CCOM,\,i}$: the match quality of cue CCOM with the item *i*
- $M_{CCOM,\,+CCOM}$: associative strength between cue CCOM and a feature *+CCOM*

[4] MASC: masculine; CCOM: c-command

The spreading activation to the target item from cue CCOM is determined by the probability of the target item given cue CCOM which is calculated as follows:

$$P(\text{target}|\text{CCOM}) = \frac{Q_{\text{CCOM, target}}}{\sum_{v \in \text{Items}} Q_{\text{CCOM, v}}}$$

As only the feature *+CCOM* of the target *actress* matches the cue CCOM of the reflexive, the match quality of cue CCOM and the feature *+CCOM* gets the maximal value of 1 (maximally associated); thus, both the numerator:

$Q_{\text{CCOM, target}} = \sum_{+\text{CCOM} \in K_{\text{target}}} M_{\text{CCOM, +CCOM}} = M_{\text{CCOM, +CCOM}} = 1;$

and the denominator:

$$\sum_{v \in \text{Items}} Q_{\text{CCOM, v}} = Q_{\text{CCOM, target}} + Q_{\text{CCOM, distractor}} = 1 + 0 = 1$$

take a value of 1 (since the match of cue CCOM to the distractor *John* equals 0).

The probability of the target *actress* given cue CCOM is then calculated as:

$$P(\text{target}|\text{CCOM}) = \frac{Q_{\text{CCOM, target}}}{\sum_{v \in \text{Items}} Q_{\text{CCOM, v}}} = \frac{1}{1} = 1$$

Therefore, the spreading activation $S_{\text{CCOM, target}}$ from cue CCOM to the target item *actress* is:

$$S_{\text{CCOM, target}} = \text{MAS} + \ln[P(\text{target}|\text{CCOM})] = \text{MAS} + \ln[1] = \text{MAS}$$

with MAS indicates *maximal associative strength*.

Since the remaining cue MASC does not match any features of the target item *actress*, $S_{\text{CCOM, target}}$ is equivalent to the maximal amount of spreading activation S_{target} that the target *actress* could possibly receive:

$$S_{\text{target}} = S_{\text{CCOM, target}} = \text{MAS}$$

In other words, the target item *actress* receives MAS (maximal associative strength) of retrieval given the cue CCOM.

b. Spreading activation of distractor

Similarly, as the cue MASC only matches the feature *+MASC* of the distractor *John*, the spreading activation $S_{MASC, John}$ of the distractor item *John* equals the total amount of spreading activation $S_{distractor}$ which is equivalent to MAS.

As previously stated, both the target *actress* and distractor *John* were assigned the subject role in the experimental sentence (i.e., item prominence concerning grammatical role). Additionally, the two items also receive the similar amount of maximal associative strength given each cue of the reflexive: $S_{target} = S_{distractor} = MAS$. Therefore, if accounting for only the total amount of spreading activation, the target *actress* and the distractor *John* have equal chance in the race for retrieval. The competitive advantage between the two items is then determined by the prominence of the distractor *John* as predicted by the LV05 model, as well as the recency advantage to the reflexive that the distractor has over the target. Compared to the target *actress*, the distractor *John* is linearly closer to the reflexive *himself*. High distractor prominence and recency advantage would possibly result in the distractor's higher probability of retrieval. This finding is consistent with some earlier studies which also found higher rate of retrieval errors in the target-mismatch condition (i.e., when the target and the reflexive do not match in gender) (e.g., Patil et al., 2016; Sturt, 2003).

5.4.1.1.2. Target-match Configuration

The actress$^{+FEM}_{+CCOM}$ that

Mary$^{+FEM}_{-CCOM}$ interviewed at the awards ceremony about two years ago described

herself$\{^{FEM}_{CCOM}\}$ as an extreme workaholic.[5]

a. Spreading activation of target

- *Spreading activation of target given cue CCOM: $S_{CCOM,\ target}$*

For the target-match configuration, the target *actress* gets 100% activation from the cue CCOM, thus its spreading activation $S_{CCOM,\ target}$ matches the total amount of spreading activation S_{target}. Alternatively stated, the target *actress* gets the maximal associative strength (MAS) given the cue CCOM, as in the case of the target *actress* in the target-mismatch configuration.

- *Spreading activation of target given cue FEM: $S_{FEM,\ target}$*

As for the cue FEM, both the target *actress* and the distractor *Mary* have the gender feature *+FEM* that matches the cue. The cross association level of the cue FEM results in the activation reduction of both the target and the distractor given cue FEM. Under this circumstance, apart from the total activation from the fully matching cue CCOM, the target *actress* also receives activation from the cue FEM which spreads half of its activation to the distractor *Mary*.

Given the activation spread between items, the probability of the target *actress* given cue FEM is calculated as:

$$P(\text{target}|\text{FEM}) = \frac{Q_{\text{FEM, target}}}{\sum_{v \in \text{Items}} Q_{\text{FEM, v}}} = \frac{0.5}{1} = 0.5$$

[5] FEM: feminine; CCOM: c-command

For the target-match condition, the match quality $Q_{FEM,\ target} = 0.5$ because the cue FEM spreads 50% of its activation to the target item *actress*, and 50% to the distractor *Mary*. Since the denominator is the sum of match between cue FEM to target item *actress* ($Q_{FEM,\ target} = 0.5$) and cue FEM to distractor item *Mary* ($Q_{FEM,\ distractor} = 0.5$), it is therefore:

$$\sum_{v \in Items} Q_{FEM,\ v} = Q_{FEM,\ target} + Q_{FEM,\ distractor} = 0.5 + 0.5 = 1$$

The amount of spreading activation $S_{FEM,\ target}$ is no longer equivalent to the maximal associative strength (MAS) as in the target-mismatch configuration, but is calculated as:

$$S_{FEM,\ target} = MAS + \ln[0.5] = MAS + [-0.7] = MAS - 0.7$$

With the spreading activation $S_{CCOM,\ target}$ of the target *actress* given cue CCOM = MAS, and the spreading activation $S_{FEM,\ target}$ given cue FEM = MAS – 0.7; the total amount of spreading activation S_{target} that the target receives given both of the two cues CCOM and FEM can be calculated as:

$$S_{target} = S_{CCOM,\ target} + S_{FEM,\ target} = \sum_{CCOM \in Cues} W_{CCOM} S_{CCOM,\ target} + \sum_{FEM \in Cues} W_{FEM} S_{FEM,\ target}$$

Since the reflexive has CCOM and FEM as its two cues, the two cues are equally weighted, which makes the weight of cue CCOM: $W_{CCOM} = 0.5$, and the weight of cue FEM: $W_{FEM} = 0.5$. Therefore, the total spreading activation of the target given both cue CCOM and cue FEM is calculated as:

$$\begin{aligned} S_{target} &= 0.5 * S_{CCOM,\ target} + 0.5 * S_{FEM,\ target} \\ &= 0.5 * MAS + 0.5 * (MAS - 0.7) \\ &= 0.5 * MAS + 0.5 * MAS - 0.5 * 0.7 \\ &= MAS - 0.35 \end{aligned}$$

b. Spreading activation of distractor given cue FEM: $S_{FEM,\ distractor}$

As the distractor *Mary* also gets activation from the cue FEM, the spreading activation $S_{FEM,\ distractor}$ of the distractor given cue FEM equals the spreading activation $S_{FEM,\ target}$ of the target given cue FEM, which is equivalent to MAS – 0.7.

The distractor item does not receive activation from any other cues, thus its spreading activation $S_{FEM,\ distractor}$ is also the total activation $S_{distractor}$ that the distractor item could get.

A comparison between the total amount of spreading activation of the target (S_{target} = MAS – 0.35) and the total amount of spreading activation of the distractor ($S_{distractor}$ = MAS – 0.7) in the target-match configuration infers that the target *actress* has a stronger spreading activation than the distractor *Mary*; which explains the higher probability of correct retrieval in the target-match than in the target-mismatch condition when both S_{target} and $S_{distractor}$ = MAS.

5.4.1.2. Item Prominence

Addition to what was discussed in the previous section, since the distractors in the target-match and target-mismatch configurations (i.e., *Mary* and *John* respectively) were both assigned the subject role as the target *actress*, the distractors were expected by the ACT-R theory to maintain a high prominence status in memory. The revised LV05 model predicts a correlation between facilitatory effect and distractor prominence in the target-mismatch condition. For the target-match, high level of distractor prominence is also expected by the model to lead to facilitation, while low level of distractor prominence will result in inhibition (Engelmann et al., 2019). The prediction of facilitation in both the target-match and the target-mismatch configuration partly explains the speed up in reading time at the reflexive (i.e., segment 4) in the target-match condition of the present study. Though facilitatory effects might occur in the target-match condition, the total spreading activation of the target *actress* is still

stronger than the total spreading activation of the distractor *Mary* as previously calculated in 5.4.1.1.2. Taking into account both the predictions of distractor prominence and spreading activation – other things being equal – the probability of the target antecedent being retrieved in the target-match configuration is still higher than the target-mismatch configuration even though there might be an effect of facilitation (speed-up reading time) in the target-match.

As reviewed in section 2.1.2.3, the revised LV05 model (Engelmann et al., 2019) presented an additional prominence component p_i to the base-line activation B_i of the item i (e.g., $p_{distractor}$ – a prominence component added to the base-line activation $B_{distractor}$ of the distractor *John*), suggesting that items with higher prominence have stronger activation, and therefore will have higher probability of retrieval. The base-line activation B_i in Equation 3 (section 2.1.2.3) is updated into:

$$B_i = \ln\left(\sum_{j=1}^{n} t_j^{-d}\right) + \beta_i + p_i$$

Another component concerning the assumption of item prominence introduced by the LV05+IP+MAC model is the saliency component which is considered as a weighting of the match quality Q_{ji} of item i given cue j. With the addition of the saliency component $\frac{1}{1+ qe^{-(B_i- \tau)}}$, the Q_{ji} is calculated as follows (with τ is the threshold of retrieval, q is the scaling constant that scales the strength of association between the saliency of an item and the match quality Q_{ji}):

$$Q_{ji} = \sum_{k \in K_i} M_{jk} \times \frac{1}{1+ qe^{-(B_i- \tau)}}$$

In this updated equation of $Q_{ji,}$ Engelmann et al. (2019) specified that when q takes a value of 0, the base-line activation B_i of the item i with prominence p_i does not appear in $P(i|j)$. When q takes values greater than 0, the associative strength between the cues and the item will

be affected by the base-line activation B_i with the added item prominence p_i. The extended LV05 model expects that the prominence of item *i* (i.e., p_i) and the association between item *i* and cue *j* (i.e., M_{ji}) will determine the retrieval probability of item *i* given cue *j* (i.e., $P(i|j)$). The addition of a salient component indicates that cues can be discriminated with varying degrees. In other words, there could be multiple features that match a cue, and that the items which are more prominent are expected to have stronger association with the cues, and thus having stronger spreading activation (Engelmann et al., 2019). This prediction of item prominence is evidently reflected in the distractor's high probability of retrieval in the target-mismatch configuration of the present study.

The computation of cue weights, spreading activation, and item prominence suggested by the updated LV05 aligns with the predictions of the ACT-R framework and the findings from a large body of earlier research on reflexive licensing. Given the consistency between the results of the current study and the findings of past research on cue-based memory retrieval, this study provides further support to the application of a cue-based access mechanism in the resolution of anaphoric reflexives by different groups of English speakers.

5.4.2. Good-enough Processing in the Resolution of Anaphoric Reflexives

In recent years, an increasing volume of research has been conducted on the application of a good-enough processing strategy in sentence processing and comprehension by native and nonnative speakers of English (e.g., Pratt & Fernández, 2016; Tan & Foltz, 2020). Karimi and Ferreira (2016) proposed the Online Cognitive Equilibrium hypothesis which was considered the impetus for the Good-enough hypothesis (Ferreira et al., 2002; Ferreira, 2003; Ferreira & Patson, 2007). The Online Cognitive Equilibrium hypothesis suggested that the language processing system generally aims for early comprehension so that the system can reach the

equilibrium state at the earliest opportunity. Once in the state of equilibrium, the system prefers staying in the state for as long as possible, or at least until the task requires otherwise. Though the question of whether or not simple heuristics and detailed syntactic algorithms are always simultaneously active or mutually exclusive during processing still remains a matter of open discussion, most findings from earlier studies have consistently pointed to the application of a resource-preserving processing route at the expense of an effortful approach to syntactic analysis, particularly when there is an increase in cognitive load. Only when required by the task, comprehenders will switch back to the syntax-based algorithmic analysis.

This study started out to test the predictions of the prosody-memory integrated model for sentence processing suggested by Pratt and Fernández (2016). It was expected that a similar pattern of results would be revealed from the current data, such that participants would resort to a good-enough processing approach when heavy cognitive load took a toll on processing. I predicted that manipulations of syntactic structures, interference effects, and implicit prosodic contours in the form of text presentation formats would result in increased cognitive load; thus, making comprehenders, regardless of their language profiles, employ a good-enough processing strategy to maintain general comprehension.

As for syntactic structures, results from the mixed model of comprehension accuracy showed that there was a marginal effect of structure (see Table 12), such that participants responded to comprehension probes of the simple sentences more accurately than the complex sentences. This finding indicated that increased syntactic complexity caused difficulty in processing among participants.

Text presentation formats also affected participants across experimental tasks. Overall, the *sentence* format which resembles the format of normal reading was associated with highest

sensitivity to grammaticality. That both groups, especially L2, performed best in the *sentence* format across experimental tasks denoted that when sentences were presented in the most "normal" fashion, the need to resort to a good-enough processing approach diminished. This finding is not in line with the findings from Pratt and Fernández's (2016) study which found low rate of grammaticality rating accuracy among participants in the *sentence* format despite the fact that the participants could provide accurate responses to the comprehension questions.

As for the other text presentation formats: *word-by-word* and *phrase-segment*, the two groups differed in their performance in the two reading paradigms. For the L1 group, the *word* format was most disruptive (though not significantly), while for L2, the *phrase-segment* format was least beneficial. Similar results were observed in Pratt and Fernández's (2016) study, signifying that once the parsers opted for a good-enough approach, none of the prosodic manipulations seemed to provide any additional help. Based on the findings, there was also an interaction between reading formats and comprehension tasks (Table 12). Participants, in general, responded to the reflexive-targeted questions most accurately in the *sentence* format, and least accurately in the *word* (for L1) and in the *phrase* format (for L2) (see Figure 11). On average, general-targeted questions had higher rate of response accuracy than reflexive-targeted questions, which suggests that fewer resources were allocated to detailed syntactic algorithm. These findings were reinforced by the mismatch effect (high probability of misretrieval) which was observed particularly across the *word* and the *phrase* formats.

The results of the current study provide additional support for the application of a good-enough strategy in sentence processing, which aligns with the main findings of Pratt and Fernández's (2016) research: (i) once the parsers settle on a heuristic approach to maintain general comprehension, neither reading formats was found to assist the parsers during

processing; (ii) only when mandated by the task, effortful syntactic analysis is applied. These findings are also in line with the Online Cognitive Equilibrium hypothesis (Karimi & Ferreira, 2016) which suggests the preference for early comprehension of the language processing system so that the equilibrium state can be reached in the shortest possible time.

The parser's application of a good-enough processing strategy during processing was further investigated via the correlation test between general comprehension and grammaticality judgment. The present study took into account the manipulations of both implicit prosody and syntactic complexity as integrated causes leading to increased cognitive load. Results from the correlation tests confirmed the trade-off effects found in Pratt and Fernández's (2016) study. The negative correlation between general comprehension and grammaticality judgment in complex constructions indicates that syntactic processing is compromised, especially under heavy cognitive load, to fulfil the task of sustaining general comprehension.

5.4.3. Implicit Prosody in the Resolution of Anaphoric Reflexives

This section discusses how manipulations of implicit prosody in the form of text presentation formats interfered with L1 and L2 processing of reflexive-antecedent dependencies.

As presented in Chapter 4, in general, participants were most facilitated by the *sentence* format across experimental tasks. The maintenance of good performance in the *sentence* condition by both of the two groups indicates that when participants were allowed to read sentences in the configuration that resembles the format of normal reading, disruption to comprehension and syntactic analysis was least likely to occur. Quite the contrary, participants in Pratt and Fernández's (2016) study did not find the *sentence* format very much facilitative during processing. The nonalignment of findings might be partly due to individual differences

among participants, and distinctive features of the two types of syntactic constructions (i.e., relative clause attachments versus reflexive-antecedent dependencies).

It was expected that the *phrase* format which segmented sentences into chunks aligning with the clausal edges of the sentence would provide guiding cues for retrieval, and thus, reducing task load. However, performance in the *phrase-segment* condition varied among tasks and between groups. Comprehension accuracy for the target-match sentences was averagely highest in the *phrase* format. However, the *phrase-segment* paradigm did not seem to help with the processing of target-mismatch sentences (particularly target-mismatch sentences in complex condition), which was reflected in the averagely lowest percentage of grammaticality rating accuracy. L1 participants generally made good use of the *phrase* format especially in the comprehension task, while L2, on the other hand, had the lowest rate of comprehension accuracy in the *phrase* condition (Figure 10). Pratt and Fernández (2016) came up with similar results concerning the adverse effects of the *phrase* format in L2 processing of relative clauses. Based on the given findings, task demands, individual differences, and implicit prosody should be treated as covariates in sentence processing and comprehension.

In general, it was predicted that participants would find reading sentences in the *word-by-word* format most disruptive in the comprehension task. However, taking into account the manipulation of syntactic complexity, participants' task performance in the *word* format was not found to differ from the other two reading formats. In fact, response accuracy to comprehension questions in the *word* format was averagely higher than in the *phrase* format (Figure 10). This finding might be attributed to the performance of L2 participants in the *word* format which was marginally better than their performance in the *phrase* condition. Pratt and Fernández (2016) suggested that the difference in L2 participants' performance in the *phrase* format as compared

to L1 might be associated with reading fluency development. Since the L2 participants of the present study varied in English proficiency, participants with lower proficiency were assumed to benefit more from the *word-by-word* paradigm (Cromer, 1970; Perfetti, 1988).

The reading times of the experimental sentences in the *sentence* format revealed a significant effect of structure, i.e., participants spent longer time reading complex sentences than simple sentences. The inclusion of a past participle clause in the complex configuration in place of the adverbial phrase in the simple configuration possibly mandated longer reading times of complex sentences. There was an effect of English reading proficiency across reading times of segments as well as of sentences, which indicated that participants with higher English reading proficiency read individual segments and sentences faster than participants with lower proficiency. The effects of reading proficiency on processing will be further discussed in the following section.

5.4.4. English Reading Proficiency in the Resolution of Anaphoric Reflexives

As specified in Chapter 3, participants were tested for reading proficiency in English following the reading experiment. I expected that the administration of a real time reading proficiency test, to some extent, would reflect participants' real time cognitive processing of sentences. In addition to the reading test, participants were also asked to complete a self-rated proficiency questionnaire. The results of the correlation test between the reading proficiency test and the self-rated questionnaire showed that there is a significant positive correlation between the two measures. However, it is noteworthy to point out that up to date, reading proficiency tests specifically dedicated to real-time processing have been very limited in quantity, if not underdeveloped, especially for languages other than English. In addition, most test developers offer limited access to the available tests. Proficiency measures in real time processing in most

earlier studies were in the form of (adapted) language proficiency tests (e.g., Hopp, 2006; Jackson, 2008; Pratt & Fernández, 2016). The absence of open-access reading proficiency measures that could be used to investigate the association between proficiency and real-time processing necessitates the need for the development of such a proficiency test. The results of the test could then be used to get insight into how proficiency – as a parameter – might affect sentence processing and comprehension. Discussion on how that could contribute to future research will be discussed in Chapter 6.

Based on the results, in general, higher English reading proficiency was associated with higher grammaticality rating accuracy and shorter reading times of individual segments and sentences. However, taking into account L1 and L2 differences, as discussed in the previous section, L2 participants did not seem to benefit from the *phrase-segment* format as much as L1 participants did across experimental tasks, which in part could be explained by L2's (especially low proficient L2) inability to immediately make use of phrase-structure information during processing (e.g., Felser, Roberts, Gross, & Marinis, 2003). As for utility of prosodic cues, under certain circumstances, L2 participants could make use of prosodic cues in retrieval of antecedents. However, when cognitive load increases as in cases of syntactic complexity or mismatch interference, effective deployment and integration of prosodic cues at the time of retrieval might be absent among low proficient L2 speakers (Fultz, 2009).

The ability to accurately distinguish between grammatical (target-match) and ungrammatical (target-mismatch) sentences among participants with higher proficiency showed that L1 and advanced L2 speakers were qualitatively similar in syntactic parsing strategies. L2 participants with lower proficiency, on the other hand, might be less automatic in the processing of morphosyntactic structures (Hoover & Dwivedi, 1998), which resulted in their less accurate

rating of grammaticality. The effects of L2 proficiency during processing were also reported in some other earlier studies where immediate sensitivity to morphosyntactic information was detected more frequently among advanced L2 speakers as compared to L2 with lower proficiency (e.g., Frenck-Mestre, 2002).

Chapter 6. Conclusion and Recommendations

The study has looked into the processing of reflexive-antecedent dependencies with respect to a prosody-memory integrated model (Pratt & Fernández, 2016) that takes into account implicit prosody, cue-based memory retrieval mechanism, and good-enough processing. The model was first introduced and tested using relative clause attachments, focusing on long-distance subject-verb agreement. I extended the predictions of the model to reflexive-antecedent dependencies of which syntactic configuration requires different types of syntactic binding. Deployment of retrieval cues during processing of reflexive anaphors was expected to follow the architecture of cue-based memory access. Nevertheless, interference effects in reflexive-antecedent dependencies are distinctive to attraction effects in subject-verb agreement, thus I predicted that this study would generate a similar pattern of findings as Pratt and Fernández's, but the syntactic as well as morphological features peculiar to reflexive licensing would not point to a completely identical processing mechanism as that of subject-verb agreement. Retrieval of cues was affected by interference effects, which inevitably had detrimental effects on the parser's computation of reflexive-antecedent dependencies.

The results are in support of a cue-based memory retrieval mechanism (Lewis & Vasishth, 2005; Engelmann et al., 2019) in the resolution of reflexive-antecedent dependencies. Interference effects in the target-match (grammatical) and target-mismatch (ungrammatical) configurations were in line with findings from most earlier research, such that higher rate of misretrieval was found in the target-mismatch condition as the result of recency and distractor prominence, and increased reading times at the reflexive were associated with the mismatch between the reflexive and the accessible target. Comprehension accuracy remained high for general comprehension despite manipulations of syntactic complexity and grammaticality

(target-match/target-mismatch configuration), while the ability to distinguish between grammatical and ungrammatical constructions decreased as interference effects and syntactic complexity increased. The findings give substantial weight to the integration of cue-based memory retrieval mechanisms into the framework of Good-enough Processing and Online Cognitive Equilibrium hypotheses. Though simple heuristics as well as syntactic algorithms might be simultaneously active during processing (Karimi & Ferreira, 2016), a resource-preserving processing route was still prioritized over a syntax-based algorithmic route for the maintenance of general comprehension, especially when detailed syntactic analysis was not required by the task or assumed to be necessary by the parser.

Prosodic contours aligning with syntactic parsing were expected to ease processing by reducing cognitive load. However, only native speakers were found to be facilitated by the text presentation format in which phrasal segments were in compliance with the sentence's clausal edges. When participants read sentences in the *sentence* format, their performance across experimental tasks exceled, indicating that syntactic and prosodic parsing were still employed during processing in the presentation format that most resembles the format of normal reading. Though the segmentation of texts in the *phrase* format resonated with the syntactic parsing of the sentence, reading fluency and memory decay might also affect processing, especially when the design of the experiment forced the preceding phrasal segment to dismiss from the computer screen in exchange for the appearance of the next segment. Reading fluency and individual differences might contribute to the varying effects of prosodic contours on the processing of anaphoric reflexives.

Additionally, as pointed out in section 5.4.4, the development of a reading proficiency test that is closely tied to the architecture of real time processing is much needed so that future

research will have a more reliable correlation measure between language proficiency and language processing. More sensitive techniques such as eye-tracking and ERPs are also recommended for better justifications of the model's predictions. As suggested by Pratt and Fernández (2016), future research might proceed from existing work by examining the application of the model in the processing of other types of syntactic configurations such as licensing of English cases, negative polarity items, etc.

As stated at the beginning of this Chapter, by investigating L1 and L2 processing of reflexive-antecedent dependencies, this study provided further support for the cognitive model that integrated implicit prosody and cue-based memory retrieval into the framework of Good-enough Processing and Online Cognitive Equilibrium hypotheses (Pratt & Fernández, 2016). However, neither this present study nor Pratt and Fernández's thoroughly investigated individual differences such as working memory capacity and developmental stages of reading fluency in the application of the model. For reflexive-antecedent dependencies, working memory capacity was suggested to modulate the associative strength between cues and the features of an item (Engelmann, et al., 2015). English reflexives have two features that typically manifest as cues during retrieval: *number* and *gender*. It is predicted that cognitive effort increases as the one-to-one association between cues and item features increases. Thus, for parsers with low working memory capacity, cross cue-feature associations and similarity-based interference effects are likely to be stronger, which consequently taxes memory, leading to inhibitory interference effects in the target-mismatch configuration. As participants in the present study were found to have higher erroneous retrieval and longer retrieval latency in the target-mismatch condition, insights into participants' working memory capacity might have

provided more plausible explanations as to the occurrence of inhibitory effects in the target-mismatch configuration.

Additionally, for the memory mechanism underlying sentence processing, there has been an ongoing debate concerning the co-occurrence of encoding interference and retrieval interference during processing (e.g., Chanales, Dudukovic, & Richte, 2019; Goh & Lu, 2012; Jager et al., 2015; Villata, Tabor, & Franck, 2018). At one end of the spectrum is the argument that similarity-based interference was found at encoding as well as at retrieval (e.g., Chanales et al., 2019; Villata et al., 2018); while at the other end, it was argued that encoding interference was not the interference effect observed at retrieval (e.g., Jager et al., 2015). The unsettling disagreement regarding encoding interference during processing motivates future studies to examine whether encoding interference and retrieval interference should be "teased apart" or integrated into one single memory mechanism. Findings from future research will possibly open a new venue for the development of a comprehensive cognitive model which could be applied by all language users, irrespective of their language backgrounds, in the processing of varied syntactic configurations.

Appendix A

Language experience and proficiency questionnaire: L1 participants

Adapted from LEAP-Q (Marian, Blumenfeld, & Kaushanskaya, 2007)

My name is Nguyen Thuy Duong - a teacher from the English Department - Hanoi University. I am now doing a Ph.D. in Psycholinguistics in the Department of Linguistics and Comparative Cultural Studies - Ca' Foscari University of Venice.

My research focuses on the processing of sentences by different groups of English speakers. In order for me to proceed with the other stages of my research, I would like to ask you to help me by answering the following questions concerning your language experience and proficiency. This survey was created to better understand the profiles of native speakers of English.

The survey will take about 15 minutes to complete. Please answer every question and give your answers sincerely. Your responses will only be used for the research purpose and your identity will never be disclosed.

***NOTE**: Following this survey is a 45-minute experiment on sentence processing which will be administered upon our agreement on a meeting schedule.*

Thank you very much for your time and participation.

I. BIOGRAPHICAL INFORMATION

Name _____

Date _____

Age_____

Gender: Male / Female / Other

Current place of residence: City/province_____
Country_____

Country of origin: _____

If your country of origin is different than your country of residence, when did you move to the country where you currently live?_____

Highest level of formal education (your current or most recent education level, even if you have not finished the degree).

- Graduate school (PhD/MD/JD)
- Graduate school (MA/MS)
- College (BA/BS)
- High school
- Other (specify):

II. LANGUAGE USE

1. Please list all the languages you know in order of dominance (your most dominant language first). If you are equally dominant in two languages, please pick an order for them.

1)

2)

2. Please list all the languages you know in order of acquisition (your native language first):

1)

2)

3. Please list what percentage of the time you are **currently** and **on average** exposed to each language. ***NOTE****: Your percentages should add up to 100%*

Language(s):		
Percentage:		

4. When choosing to **read** a text available in all your languages, in what percentage of cases would you choose to read it in each of your languages? Assume that the original was written in another language, which is unknown to you. ***NOTE****: Your percentages should add up to 100%*

Language(s):		
Percentage:		

5. When choosing a language to SPEAK with a person who is equally fluent in all your languages, what percentage of time would you choose to speak each language? Please report percent of total time. ***NOTE****: Your percentages should add up to 100%*

Language(s):		
Percentage:		

6. Please name the cultures with which you identify. On a scale from 0 to 10, please rate the extent to which you identify with each culture. (*E.g.: English: 10; Vietnamese: 8*).

III. ENGLISH LANGUAGE: *All questions below refer to English - your native language.*

7. Age when you…

began acquiring English:	became fluent in English:	began reading in English:	became fluent reading in English:

8. Please list the number of years and months you spent in each language environment:

	Years	Months
A country/countries where English is spoken		
A family/families where English is spoken		
A school and/or working environment where English is spoken		

9. On a scale from 0 to 10, please select your level of proficiency in…

0	1	2	3	4	5	6	7	8	9	10
none	very low	low	fair	slightly less than adequate	adequate	slightly more than adequate	good	very good	excellent	perfect

Speaking	
Understanding spoken langu	
Reading	

10. On a scale from 0 to 10, please select how much the following factors contributed to your using English:

0	1	2	3	4	5	6	7	8	9	10
not a contributor					moderate contributor					most important contributor

Interacting with friends	
Interacting with family	
Reading	
Writing for school/work	

11. Please rate to what extent you are currently exposed to English in the following contexts:

0	1	2	3	4	5	6	7	8	9	10
never	almost never				half of the time					always

Interacting with friends	
Interacting with family	
Watching TV	
Listening to radio/podcasts	
Listening to music	
Reading for fun	
Reading for school/work	

12. Please rate how frequently others identify you as a native speaker of English:

0	1	2	3	4	5	6	7	8	9	10
never	almost never				half of the time					always

Appendix B

Language profile questionnaire: L2 participants

I. BIOGRAPHICAL INFORMATION

Name _____

Date _____

Age_____

Gender: Male / Female / Other

Current place of residence: City/province_____
Country_____

Country of origin: _____

If your country of origin is different than your country of residence, when did you move to the country where you currently live?_____

Highest level of formal education (your current or most recent education level, even if you have not finished the degree).

- Graduate school (PhD/MD/JD)
- Graduate school (MA/MS)
- College (BA/BS)
- Other (specify):

II. LANGUAGE HISTORY: *In this section, I would like you to answer some factual questions about your language history.*

1. Please list all the languages you know in order of dominance (your most dominant language first). If you are equally dominant in two languages, please pick an order for them.

1)

2)

3)

4)

2. At what age did you **start learning** English? (Please circle)

Since birth 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+

3. At what age did you **start to feel comfortable** using English? (Please circle)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+ Not yet Can't remember

4. How many **years** of **English language classes** have you had (primary school through university)? (Please circle)

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+

5. How many **years** of **classes (grammar, history, math, etc.)** have you had in English (primary school through university)? (Please circle)

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+

6. Please indicate the age at which you **started using English** in each of the following environments.

At home	With friends	At school	At work	Language learning software	Online games	Social media

7. If you have lived or travelled in countries where you used English for **three or more months**, please indicate the name of the country, the length of your stay, and how often you used English for each country, using the following scale.

Never	Rarely	Sometimes	Regularly	Usually	Always
1	2	3	4	5	6

** You may have been to the country on multiple occasions, each for a different length of time. Add all the trips together.*

Country	Length of stay [month(s)]	Frequency of use (1 – 6)

8. How much time have you spent in a **family** where English was spoken?

Indicate months or years:

9. How much time have you spent in a **work or school environment** where English is spoken?

Indicate months or years:

III. LANGUAGE USE: *In this section, I would like you to answer some questions about your language use.*

10. In an average week, how many hours do you use English **with friends**?

Indicate hours:

11. In an average week, how many hours do you use English **with family**?

Indicate hours:

12. In an average week, how many hours do you use English **at school/work**?

Indicate hours:

13. Please estimate the number of hours per week that you are **exposed to English.**

Indicate hours:

14. Please estimate how many hours per week you spend speaking English with the people listed below:

	Hours per week
Family members	
Friends *(Include significant others if you did not include them as family members. E.g., married partners)*	
Classmates	
Coworkers *(Include anyone in the work environment. I.e., if you are a teacher, include students as coworkers)*	
People on the Internet	

15. How often do you use English for the following activities? Please enter the number in the table according to the scale below:

Never	Rarely	Sometimes	Regularly	Usually	Always
1	2	3	4	5	6

	Frequency
Thinking	
Talking to yourself	
Expressing Emotion *(including shouting, cursing, showing affection, etc.)*	
Arithmetic *(including counting, calculating tips, etc.)*	
Remembering numbers *(including telephone numbers, ID numbers, etc.)*	

16. Please estimate how many hours per week you spend engaged in the following activities in English.

	Hours per week
Watching television	
Listening to music	
Listening to radio or podcasts	
Reading for fun	
Reading for school/work	
Chatting on social media	
Texting/Writing emails	
Writing for school/work	

IV. LANGUAGE PROFICIENCY: *In this section, I would like you to rate your language proficiency.*

Not well at all = 0

Very well = 6

17. How well do you **speak English**? (Please circle)

0 1 2 3 4 5 6

18. How well do you **understand English**? (Please circle)

0 1 2 3 4 5 6

19. How well do you **read English**? (Please circle)

0 1 2 3 4 5 6

20. How well do you **write English**? (Please circle)

0 1 2 3 4 5 6

21. Using the CEFR, what would you self-rate your **level of English,** whether or not you have a certification? (Please circle)

A1 A2 B1 B2 C1 C2

22. If you have taken any **standardized language proficiency tests** (e.g., TOEFL, IELTS, PET), please write the name of each test and the score you received. If you do not remember the exact score, then indicate an "Approximate score" instead.

Test	Score	(Approximate score)	Date Taken

V. LANGUAGE ATTITUDES: *In this section, I would like you to respond to statements about language attitudes.*

Disagree = 0

Agree = 6

23. I feel like myself when I speak English. (Please circle)

0 1 2 3 4 5 6

24. I identify myself with an English-speaking culture. (Please circle)

0 1 2 3 4 5 6

25. It is important to me to use (or eventually use) English like a native speaker. (Please circle)

0 1 2 3 4 5 6

26. I want others to think I am a native speaker of English. (Please circle)

0 1 2 3 4 5 6

27. Please enter the language you feel the **most comfortable** in when listening, speaking, reading, and writing in each of the contexts listed below:

	Listening	Speaking	Reading	Writing
At home				
With friends				
At school				
At work				
On the Internet				
On social media				

28. Please rate your **language learning skill**. In other words, how good do you feel you are at learning new languages, relative to your friends or other people you know? (Please circle)

Very poor	Poor	Limited	Average	Good	Very good
1	2	3	4	5	6

29. Please comment below to indicate any additional answers to any of the questions above that you feel better describe your language background or usage.

Thank you very much for your participation.

Appendix C

Experimental items

a. Simple, Grammatical
b. Simple, Ungrammatical
c. Complex, Grammatical
d. Complex, Ungrammatical

1. a. The actress that Mary interviewed at the awards ceremony about two years ago described herself as an extreme workaholic.
   b. The actress that John interviewed at the awards ceremony about two years ago described himself as an extreme workaholic.
   c. The actress that Mary interviewed at the awards ceremony held outside the theater described herself as an extreme workaholic.
   d. The actress that John interviewed at the awards ceremony held outside the theater described himself as an extreme workaholic.
2. a. The man that Richard recommended for the new job in the prestigious company convinced himself to take the job.
   b. The man that Katie recommended for the new job in the prestigious company convinced herself to take the job.
   c. The man that Richard recommended for the prestigious job offered by Louis Vuitton convinced himself to take the job.
   d. The man that Katie recommended for the prestigious job offered by Louis Vuitton convinced herself to take the job.
3. a. The girl that Sophie accompanied to the auditorium of the new campus eventually found herself a place to sit.
   b. The girl that Thomas accompanied to the auditorium of the new campus eventually found himself a place to sit.
   c. The girl that Sophie accompanied to the auditorium filled with 500 students eventually found herself a place to sit.
   d. The girl that Thomas accompanied to the auditorium filled with 500 students eventually found himself a place to sit.
4. a. The spokeswoman that Linda met at the international summit a year ago today represented herself as the chief delegate.
   b. The spokeswoman that Jason met at the international summit a year ago today represented himself as the chief delegate.
   c. The spokeswoman that Linda met at the international summit organized in early June represented herself as the chief delegate.
   d. The spokeswoman that Jason met at the international summit organized in early June represented himself as the chief delegate.
5. a. The businesswoman that Anna consulted about the start-up projects for young entrepreneurs introduced herself as chief manager of sales.
   b. The businesswoman that Tony consulted about the start-up projects for young entrepreneurs introduced himself as chief manager of sales.
   c. The businesswoman that Anna consulted about the start-up projects implemented by new graduates introduced herself as manager of sales.
   d. The businesswoman that Tony consulted about the start-up projects implemented by new graduates introduced himself as manager of sales.

6. a. The saleswoman that Carol met at the off-price department store two days ago promised herself not to work late.
b. The saleswoman that Leo met at the off-price department store two days ago promised himself not to work late.
c. The saleswoman that Carol met at the off-price department store located near Macy's promised herself not to work late.
d. The saleswoman that Leo saw at the off-price department store located near Macy's promised himself not to work late.

7. a. The schoolboy that Peter studied with in the science class back at high school finally got himself to Harvard.
b. The schoolboy that Nancy studied with in the science class back at high school finally got herself to Harvard.
c. The schoolboy that Peter studied with in the science class organized for honors students finally got himself to Harvard.
d. The schoolboy that Nancy studied with in the science class organized for honors students finally got herself to Harvard.

8. a. The lady that Sarah helped cross the road during yesterday's high tide injured herself after falling into the waters.
b. The lady that Liam helped cross the road during yesterday's high tide injured himself after falling into the waters.
c. The lady that Sarah helped cross the road flooded by high tide injured herself after falling into the waters.
d. The lady that Liam helped cross the road flooded by high tide injured himself after falling into the waters.

9. a. The butcherboy that Henry spotted at the crime scene around eight a.m. this morning convinced himself to remain silent.
b. The butcherboy that Janet spotted at the crime scene around eight a.m. this morning convinced herself to remain silent.
c. The butcherboy that Henry spotted at the crime scene surrounded by the yellow tape convinced himself to remain silent.
d. The butcherboy that Janet spotted at the crime scene surrounded by the yellow tape convinced herself to remain silent.

10. a. The schoolgirl that Cathy met at the library every Friday after class familiarized herself with the new cafeteria.
b. The schoolgirl that Brain met at the library every Friday after class familiarized himself with the new cafeteria.
c. The schoolgirl that Cathy met at the library located near the main building familiarized herself with the new cafeteria.
d. The schoolgirl that Brian met at the library located near the main building familiarized himself with the new cafeteria.

11. a. The girl that Kylie helped in the large-scale internship project from time to time reminded herself to try harder.
b. The girl that Jimmy helped in the large-scale internship project from time to time reminded himself to try harder.

c. The girl that Kylie helped in the large-scale internship project initiated two years ago reminded herself to try harder.
d. The girl that Jimmy helped in the large-scale internship project initiated two years ago reminded himself to try harder.

12. a. The weathergirl that Chloe met at the TV station two or three years ago criticized herself for being late.
b. The weathergirl that Kevin met at the TV station two or three years ago criticized himself for being late.
c. The weathergirl that Chloe met at the TV station run by a large corporation criticized herself for being late.
d. The weathergirl that Kevin met at the TV station run by a large corporation criticized himself for being late.

13. a. The choirgirl that Scarlett recruited in the singing group of ten talented members promised herself to practice all night.
b. The choirgirl that Jackson recruited in the singing group of ten talented members promised himself to practice all night.
c. The choirgirl that Scarlett recruited in the singing group led by a vocalist promised herself to practice all night.
d. The choirgirl that Jackson recruited in the singing group led by a vocalist promised himself to practice all night.

14. a. The guy that George trained at the fitness center of the luxury residential complex injured himself in a workout.
b. The guy that Emma trained at the fitness center of the luxury residential complex injured herself in a workout.
c. The guy that George trained at the fitness center operated by an athletic company injured himself in a workout.
d. The guy that Emma trained at the fitness center operated by an athletic company injured herself in a workout.

15. a. The man that Sean saw during the family reunion at the old Italian restaurant drove himself home at night.
b. The man that Denise saw during the family reunion at the old Italian restaurant drove herself home at night.
c. The man that Sean saw during the family reunion organized about two weeks ago drove himself home at night.
d. The man that Denise saw during the family reunion organized about two weeks ago drove herself home at night.

16. a. The woman that Susan met at a department store in New York occasionally treated herself to a shopping spree.
b. The woman that David met at a department store in New York occasionally treated himself to a shopping spree.
c. The woman that Susan met at a department store built long ago occasionally treated herself to a shopping spree.
d. The woman that David met at a department store built long ago occasionally treated himself to a shopping spree.

17. a. The gentleman that Eric suggested to appear on a TV show last Monday drove himself to the studio early.

b. The gentleman that Rose suggested to appear on a TV show last Monday drove herself to the studio early.
c. The gentleman that Eric suggested to appear on a show broadcast on Monday drove himself to the studio early.
d. The gentleman that Rose suggested to appear on a show broadcast on Monday drove herself to the studio early.

18. a. The gentleman that James nominated to the advisory group of the children's hospital introduced himself to all the doctors.
b. The gentleman that Ella nominated to the advisory group of the children's hospital introduced herself to all the doctors.
c. The gentleman that James nominated to the advisory group coordinated by senior professionals introduced himself to all the doctors.
d. The gentleman that Ella nominated to the advisory group coordinated by senior professionals introduced herself to all the doctors.

19. a. The lady that Laura crashed into on the road to the five-star resort miraculously protected herself from getting hurt.
b. The lady that Logan crashed into on the road to the five-star resort miraculously protected himself from getting hurt.
c. The lady that Laura crashed into on the road blocked by fallen trees miraculously protected herself from getting hurt.
d. The lady that Logan crashed into on the road blocked by fallen trees miraculously protected himself from getting hurt.

20. a. The landlady that Lily met outside the main building of the residential area reminded herself to collect the rent.
b. The landlady that Adam met outside the main building of the residential area reminded himself to collect the rent.
c. The landlady that Lily met outside the main building covered by metallic panels reminded herself to collect the rent.
d. The landlady that Adam met outside the main building covered by metallic panels reminded himself to collect the rent.

21. a. The father that Ethan invited to the parent-teacher meeting last Sunday morning introduced himself to all the other parents.
b. The father that Hannah invited to the parent-teacher meeting last Sunday morning introduced herself to all the other parents.
c. The father that Ethan invited to the parent-teacher meeting scheduled on Sunday introduced himself to all the other parents.
d. The father that Hannah invited to the parent-teacher meeting scheduled on Sunday introduced herself to all the other parents.

22. a. The mother that Natalie met at the supermarket earlier this morning taught herself how to cook by studying cookbooks.
b. The mother that Michael met at the supermarket earlier this morning taught himself how to cook by studying cookbooks.
c. The mother that Natalie met at the supermarket situated in Southampton taught herself how to cook by studying cookbooks.
d. The mother that Michael met at the supermarket situated in Southampton taught himself how to cook by studying cookbooks.

23. a. The grandfather that Owen met at the old folk's home almost four years ago unfortunately hurt himself by accident.
b. The grandfather that Bella met at the old folk's home almost four years ago unfortunately hurt herself by accident.
c. The grandfather that Owen met at the old folk's home funded by the Government unfortunately hurt himself by accident.
d. The grandfather that Bella met at the old folk's home funded by the Government unfortunately hurt herself by accident.

24. a. The lady that Lucy greeted at the lobby of the first-class hotel often treated herself very luxurious hotel stays.
b. The lady that Isaac greeted at the lobby of the first-class hotel often treated himself very luxurious hotel stays.
c. The lady that Lucy greeted at the lobby designed in the Deco style often treated herself luxurious hotel stays.
d. The lady that Isaac greeted at the lobby designed in the Deco style often treated himself luxurious hotel stays.

25. a. The spokesman that Andrew met at the conference yesterday morning introduced himself as the representative of the consultancy firm.
b. The spokesman that Elena met at the conference yesterday morning introduced herself as the representative of the consultancy firm.
c. The spokesman that Andrew met at the conference organized yesterday introduced himself as the representative of the consultancy firm.
d. The spokesman that Elena met at the conference organized yesterday introduced herself as the representative of the consultancy firm.

26. a. The boy that Joshua competed with at the national chess championship last year registered himself in this year's competition.
b. The boy that Eva competed with at the national chess championship last year registered herself in this year's competition.
c. The boy that Joshua competed with at the national chess championship organized annually registered himself in this year's competition.
d. The boy that Eva competed with at the national chess championship organized annually registered herself in this year's competition.

27. a. The aunt that Helen visited at the isolated farm of 200 hectares convinced herself that the place was safe.
b. The aunt that Ryan visited at the isolated farm of 200 hectares convinced himself that the place was safe.
c. The aunt that Helen visited at the isolated 200 hectare farm constantly convinced herself that the place was safe.
d. The aunt that Ryan visited at the isolated 200 hectare farm constantly convinced himself that the place was safe.

28. a. The uncle that Matthew called this morning on the train to the airport prepared himself well for the operation.
b. The uncle that Jolie called this morning on the train to the airport prepared herself well for the operation.
c. The uncle that Matthew called this morning on the train bound for London prepared himself well for the operation.

d. The uncle that Jolie called this morning on the train bound for London prepared herself well for the operation.

29. a. The little boy that Bill babysat at the large red-brick house every Tuesday night occasionally got himself into trouble.
b. The little boy that Julia babysat at the large red-brick house every Tuesday night occasionally got herself into trouble.
c. The little boy that Bill babysat at the large red-brick house situated in Newport occasionally got himself into trouble.
d. The little boy that Julia babysat at the large red-brick house situated in Newport occasionally got herself into trouble.

30. a. The niece that Ashley drove to the theme park with other two toddlers amused herself with the Ferris-wheel ride.
b. The niece that Aaron drove to the theme park with other two toddlers amused himself with the Ferris-wheel ride.
c. The niece that Ashley drove to the theme park renovated earlier this year amused herself with the Ferris-wheel ride.
d. The niece that Aaron drove to the theme park renovated earlier this year amused himself with the Ferris-wheel ride.

31. a. The nun that Vivian met at the charity event not very long ago committed herself to helping homeless children.
b. The nun that Jayden met at the charity event not very long ago committed himself to helping homeless children.
c. The nun that Vivian met at the charity event publicized on social media committed herself to helping homeless children.
d. The nun that Jayden met at the charity event publicized on social media committed himself to helping homeless children.

32. a. The priest that Anthony helped with the fundraising activities from time to time stopped himself from appearing in public.
b. The priest that Maria helped with the fundraising activities from time to time stopped herself from appearing in public.
c. The priest that Anthony helped with the fundraising activities every now and then stopped himself from appearing in public.
d. The priest that Maria helped with the fundraising activities every now and then stopped herself from appearing in public.

33. a. The waitress that Margaret often met at the restaurant of the hotel contented herself with a bowl of soup.
b. The waitress that Jonathan often met at the restaurant of the hotel contented himself with a bowl of soup.
c. The waitress that Margaret met at the restaurant of the 5-star hotel contented herself with a bowl of soup.
d. The waitress that Jonathan met at the restaurant of the 5-star hotel contented himself with a bowl of soup.

34. a. The ballerina that Isabel encouraged to audition for the leading role in the show found herself lonely at work.
b. The ballerina that Robert encouraged to audition for the leading role in the show found himself lonely at work.

c. The ballerina that Isabel encouraged to audition for the show choreographed by Russian artists found herself lonely at work.
d. The ballerina that Robert encouraged to audition for the show choreographed by Russian artists found himself lonely at work.

35. a. The heiress that Daisy gossiped with at the party in the exclusive nightclub assured herself that everything was perfect.
b. The heiress that Daniel gossiped with at the party in the exclusive nightclub assured himself that everything was perfect.
c. The heiress that Daisy gossiped with at the party organized by the mayor assured herself that everything was perfect.
d. The heiress that Daniel gossiped with at the party organized by the mayor assured himself that everything was perfect.

36. a. The policewoman that Lucia assisted on a case earlier this year repeatedly put herself in danger to help others.
b. The policewoman that Brandon assisted on a case earlier this year repeatedly put himself in danger to help others..
c. The policewoman that Lucia assisted on a case led by senior officers put herself in danger to help others.
d. The policewoman that Brandon assisted on a case led by senior officers put himself in danger to help others.

37. a. The businessman that Charles requested to meet at yesterday's conference on certain occasions forced himself to work all night.
b. The businessman that Elizabeth requested to meet at yesterday's conference on certain occasions forced herself to work all night.
c. The businessman that Charles requested to meet at the conference organized yesterday often forced himself to work all night.
d. The businessman that Elizabeth requested to meet at the conference organized yesterday often forced herself to work all night.

38. a. The chairwoman that Juliette mistook for another person at the ceremony the other night separated herself from the crowd.
b. The chairwoman that Mark mistook for another person at the ceremony the other night separated himself from the crowd.
c. The chairwoman that Juliette mistook for another person at the ceremony held last night separated herself from the crowd.
d. The chairwoman that Mark mistook for another person at the ceremony held last night separated himself from the crowd.

39. a. The bride that Rachel helped with the wedding ceremony of 200 guests prepared herself well for the big day.
b. The bride that Steven helped with the wedding ceremony of 200 guests prepared himself well for the big day.
c. The bride that Rachel helped with the ceremony held on the beach prepared herself well for the big day.
d. The bride that Steven helped with the ceremony held on the beach prepared himself well for the big day.

40. a. The groom that Carlos worked with at the advertising company a decade ago familiarized himself with the pre-wedding rituals.

b. The groom that Amy worked with at the advertising company a decade ago familiarized herself with the pre-wedding rituals.
c. The groom that Carlos worked with at the company located on Court street familiarized himself with the pre-wedding rituals.
d. The groom that Amy worked with at the company located on Court street familiarized herself with the pre-wedding rituals.

41. a. The bridesmaid that Joanna asked to hold the bouquet of 20 red roses poured herself a glass of wine.
b. The bridesmaid that Edward asked to hold the bouquet of 20 red roses poured himself a glass of wine.
c. The bridesmaid that Joanna asked to hold the bouquet tied with gold ribbons poured herself a glass of wine.
d. The bridesmaid that Edward asked to hold the bouquet tied with gold ribbons poured himself a glass of wine.

42. a. The actress that Elsa saw at the film festival in Venice last week denied herself a new movie offer.
b. The actress that Victor saw at the film festival in Venice last week denied himself a new movie offer.
c. The actress that Elsa saw at the international film festival organized in Venice denied herself a new movie offer.
d. The actress that Victor saw at the international film festival organized in Venice denied himself a new movie offer.

43. a. The princess that Nicole interviewed at the royal state banquet about five days ago enjoyed herself at the event.
b. The princess that Patrick interviewed at the royal state banquet about five days ago enjoyed himself at the event.
c. The princess that Nicole interviewed at the banquet hosted yesterday by the royal family enjoyed herself at the event.
d. The princess that Patrick interviewed at the banquet hosted yesterday by the royal family enjoyed himself at the event.

44. a. The old man that Frank took to the hospital at around 8:30 this morning seriously injured himself by accident.
b. The old man that Lola took to the hospital at around 8:30 this morning seriously injured herself by accident.
c. The old man that Frank took to the hospital located near the city center seriously injured himself by accident.
d. The old man that Lola took to the hospital located near the city center seriously injured herself by accident.

45. a. The witch that Selena spotted near the dark forest from time to time dragged herself to a secret tunnel.
b. The witch that Marcus spotted near the dark forest from time to time dragged himself to a secret tunnel.
c. The witch that Selena spotted near the dark forest covered by powdery snow dragged herself to a secret tunnel.
d. The witch that Marcus spotted near the dark forest covered by powdery snow dragged himself to a secret tunnel.

46. a. The stewardess that Kate bumped into on the flight to Paris quickly poured herself a hot cup of coffee.
b. The stewardess that Bradley bumped into on the flight to Paris quickly poured himself a hot cup of coffee.
c. The stewardess that Kate bumped into on the flight bound to Paris poured herself a hot cup of coffee.
d. The stewardess that Bradley bumped into on the flight bound to Paris poured himself a hot cup of coffee.

47. a. The businesswoman that Jane worked with at the exporting company in the U.S found herself a perfect significant other.
b. The businesswoman that Pedro worked with at the exporting company in the U.S found himself a perfect significant other.
c. The businesswoman that Jane worked for at the exporting company established in 1990 finally found herself a significant other.
d. The businesswoman that Pedro worked for at the exporting company established in 1990 finally found himself a significant other.

48. a. The nanny that Lena hired to homeschool the three children every so often encouraged herself not to quit the job.
b. The nanny that Hugo hired to homeschool the three children every so often encouraged himself not to quit the job.
c. The nanny that Lena hired to homeschool the children adopted since infancy encouraged herself not to quit the job.
d. The nanny that Hugo hired to homeschool the children adopted since infancy encouraged himself not to quit the job.

Appendix D

Filler items

Constructions with possessive adjectives:

1. The man who worked in this company prepared his own lunch every day before heading to work at 7.
2. The bridesmaid who arrived late at the wedding ceremony brought her own food to the reception around 10 today.
3. The groom who drove a black Mercedes showed up to his own bachelor party quite late at night.
4. The woman who worked for one of the most popular fundraising organizations managed her own charity in downtown Dallas.
5. The businesswoman who graduated from the most renowned university in town established her own company almost 15 years ago.
6. The heiress who inherited massive fortunes from her father purchased her own private jet this month for $20 million.
7. The landlady who owned the most valuable estate in town sold her own apartment at a very high price.
8. The waitress who worked in the best restaurant in town wished to open her own Italian restaurant one day.
9. The actress who starred in one of the most successful movies of all time opened her own production company.
10. The policewoman who worked on this particular case trained her own sniffer dogs to detect illegal substances during investigations.
11. The princess who just returned to the country invited her own guests to the palace for the annual banquet.
12. The man who helped extinguish the fire in the neighborhood nearly lost his own home to the fierce blaze.
13. * The stewardess who worked for American Airlines developed his own regime so as to improve efficiency during every flight.
14. * The choirboy participated in the school orchestra decided to bring her own musical instrument to school for yesterday's performance.
15. * The saleswoman who worked for this famous multinational chain of beauty stores wanted to develop his own cosmetics line.
16. *The woman who got married to the celebrity chef revealed his own secrets to a reporter two weeks ago.
17. * The lady who owned a small craft shop in the city center enjoyed knitting sweaters for his own children.
18. * The grandfather who lived in the foster home often collected her own pension at the beginning of each month.
19. * The guy who worked in the sales department of this company invested his own money in stocks and bonds.
20. * The woman who just opened a lovely bakery across the street also made his own money from food photography.
21. * The boy who won the first prize in the chess contest invited her own friends to a celebration party.

22. * The schoolgirl who volunteered to help out on the school's field trip brought his own lunchbox to school everyday.
23. * The actress who refused to appear on a TV show called his own doctor to schedule a health check.
24. * The Englishman who sold handmade handbags in a store near the center provided her own customers with excellent service.

Constructions with PP-embedded:

1. The front door to the lecture hall was being repaired so students entered the hall through the back entrance.
2. The colored glass of the window in the bedroom was broken after the storm hit the city yesterday.
3. The leather cover of the book is now covered with a thick layer of dust after years of disuse.
4. The last two seasons of the TV series have received mixed reviews from both the viewers and the critics.
5. The innovative idea from the young scientist was highly evaluated by all the senior members of the judging panel.
6. The first song of the acoustic album was composed by the youngest and most talented member of the band.
7. The key to the apartment was nowhere to be found so they had to wait outside in the snow.
8. The Q&A session of the presentation was unexpectedly extended one more hour as there were lots of questions asked.
9. * The surprise gift for the new employees were shipped to the office this morning right before the weekly meeting.
10. * The picture of the students of the class 2003 were taken by the main building on a sunny day.
11. * The painting collection of the two royal families were considered one of the most valuable collections in the nation.
12. * A big bowl of ramen noodles are being served tonight as the kids requested to have noodles for dinner.
13. * The huge box of chocolate bars were given to the best student of the month as a surprise present.
14. * The white bird from the nest up in the trees were trying hard to fly over the tree branch.
15. * The tall pile of old newspapers were stacked up against the wall to make room for a new cabinet.
16. * The text message from the telecommunications service provider were not delivered to all customers due to unexpected technical problems.

Constructions with noun complement clause:

1. The reason that they gave us for their absence at the annual year-end meeting of the company is unbelievable.
2. The email that the salesperson sent to the customer does not truly reflect what had happened at the store.
3. The fact that this city is listed as the most polluted city in the world needs to be corrected.

4. The friend that came over to have dinner with her roommate last evening speaks Japanese and English very fluently.
5. The doctor that discovered the new treatment for this fatal disease speaks very humbly about himself and his achievements.
6. The suitcase that her colleague in the same Department brought along this trip has many scratches on the front.
7. The company that she has always wanted to work for is now extending its branch to other geographical regions.
8. The watermelon that they bought in one of the grocery stores in town is juicy and tastes amazingly sweet.
9. * The thought that there might be no one coming to the event this evening make him feel extremely anxious.
10. * The possibility that machines would replace human beings one day appear fascinating but pretty frightening at the same time.
11. * The cake that her classmates gave her on her birthday have five layers frosted in very delicious vanilla buttercream.
12. * The book that she put on the table have a black leather cover wrapped around with red silk ribbons.
13. * The thought that he would travel with her to their most favorite cities put a smile on his face.
14. * The corridor that they decorated with colorful light bulbs of different shapes lead to a garden of red roses.
15. * The theory that there exists the so-called multiverse inspire scientists in their search for evidence to justify the hypothesis.
16. * The hope that she will become a very successful pianist one day always bring joy to her mother's heart.

Constructions with adverbial clause:

1. When Kelly found out about the news, she immediately rushed to the company to see what had happened there.
2. They love going to this beach city in July every summer as they feel very relaxed near the ocean.
3. Last year they took the statistics course as they need to have knowledge on statistical analysis for their study.
4. During the past few months there has been an increase in the number of members enrolling in the program.
5. Last night they attended the most awaited music festival, then headed to their favorite bar to have some drinks.
6. Before the ceremony, the organizer did a quick check around the hall to make sure everything was working fine.
7. After the class, Mary stayed at school for two more hours to meet with some of her best students.
8. Before Alice realized what was going on, Tom had left her standing speechless in the middle of the street.
9. * Whenever John feels tired, he immediately pours a cat out of the sink then finds himself something to eat.

10. * After hearing some strange noise outside, Alice decided to sleep the floor hard to find out what that was.
11. * At six o'clock, the couples got up, brushed their teeth, washed their three days, then ate a big breakfast.
12. * Without any hesitation Tony accepted the job offer then happily fell the director of the company for the opportunity.
13. * When it rained outside the little girl opened suddenly then ran quickly into the house to tell her mother.
14. * Before going out Jamie decided to close so that when she returned she would not have to do it.
15. * The two friends finally got back home two hours ago after a very long fun from the camping site.
16. * The gardener has fallen a very good job so the house owner gave him a generous tip for that.

Constructions with relative clause:

1. The man who insisted on giving a speech at the summit always refuses to attend the follow-up press conference.
2. The athlete who won the gold medal almost always requests to have the prize given to her in cash.
3. The chef who bought the very famous Italian restaurant downtown desperately wants to compete in a national cooking contest.
4. The actor who signed a commercial with a famous brand often feels exhausted at the end of the day.
5. The architect who designed the interior of this villa now wants to work freelance after years of being employed.
6. The mechanic who changed the flat tires of this used car often comes to work early in the morning.
7. The singer who won the Award for best female artist wants to build a studio for herself at home.
8. The teaching assistant who helped the language teacher organize the class activities often receives positive feedback from the students.
9. The reporter who interviewed Leo at the movie premiere last evening loves his job more and more each day.
10. The producer who cast all the actors for the movie wants to have a talk with the director now.
11. The pilot who secretly had a five-year affair with one member of the crew tells lies all the time.
12. The artist who made wonderful sculptures using only aluminium foil works very hard at the studio every single day.
13. * The secretary who often prepared the documents for the director before every meeting hate what she is doing now.
14. * The salesperson who accidentally gave the customer the shirt in the wrong size often makes similar mistake at work.
15. * The student who complained about the grade of her French test never feel happy with any grades she receives.

16. * The police officer who managed to solve the difficult and complicated case often take very cautious steps in investigation.
17. * The journalist who wrote a very long and detailed article about this incident always feel highly motivated at work.
18. * The teacher who advised her students to register for this national competition teach science and math this academic year.
19. * The receptionist who just helped the old lady carry the big and heavy bag upstairs often help other people.
20. * The flight attendant who helped the young mother to calm the baby down have three children of her own.
21. * The programmer who designed this useful and practical software package always tell himself to never stop learning new things.
22. * The shop owner who decided to rebuild his souvenir shop two years ago provide good service to the customers.
23. * The book author who wrote this award-winning novel always observe real people for further ideas in her new books.
24. * The designer who opened his own shop last month sometimes also provide makeup service to celebrity customers upon request.

Constructions containing countable/uncountable nouns:

1. The children's father always reminds them to drink a lot of mineral water especially during long hot summer days.
2. The head of the Personnel Department often receives a large amount of applications during this time of the year.
3. The beautician of the salon downtown often gets lots of positive feedback from the customers for her nail artwork.
4. The wife of the family next door regularly buys a lot of new clothes from her most favorite brand.
5. The accountant of the insurance company has recently filed many reports on the company's revenue during the past months.
6. The school bus driver always stops at all the pickup points even when a student is absent that day.
7. The first chapter of the story introduces the readers to many characters to prepare them for the follow-up sections.
8. The students of the class created lots of unforgettable and beautiful memories on the very last day of school.
9. * Mary's parents always give her a lot of practical advices on how to become a good student at school.
10. * The chef of this restaurant often serves many salads as the appetizing dish of his signature five course menu.
11. * The manager of the shopping mall told his secretary to organize at least two consecutive group meeting yesterday morning.
12. * The bartender has bought lots of different type of alcohol to test his new cocktail recipes for the nightclub.
13. * The park ranger noticed that a large amount of tree had been illegally chopped down some time last week.

14. * The students expressed their wish to have more reference book added to the linguistics section of the school library.
15. * The electrician was asked to come to the apartment yesterday to fix the two light bulb in the kitchen.
16. * The Housing Office official told the two students to sign three different document before handing them the apartment keys.

Appendix E

Reading Test

(Adapted from Reading Test 1, Part 2, Cambridge Certificate in Advanced English tests, Vol. 6, Cambridge Books for Cambridge Exams, 2005)

Read the following passage silently and fill in each blank with a suitable word.

Verbally READ ALOUD the missing words while you read the passage.

Enjoy the benefits of stress!

Are you looking forward to another busy week? You should be according to some experts. They argue that the stress encountered in **(1)** ..... daily lives is not only good for us, but essential to survival. They say that the response to stress, which creates a chemical called adrenalin, helps the mind and body to act quickly **(2)** ..... emergencies. Animals and human beings use it to meet the hostile conditions **(3)** ..... exist on the planet.

Whilst nobody denies the pressures of everyday life, what is surprising is that we are yet to develop successful ways of dealing with them. **(4)** ..... the experts consider the current strategies to **(5)** ..... inadequate and often dangerous. They believe that **(6)** ..... of trying to manage our response to stress with drugs or relaxation techniques, we must exploit it. Apparently, research shows that people **(7)** ..... create conditions of stress for **(8)** ..... by doing exciting and risky sports or looking for challenges, cope much better with life's problems. Activities of this type **(9)** ..... been shown to create a lot of emotion; people may actually cry or feel extremely uncomfortable. But there is a point **(10)** ..... which they realise they have succeeded and know that it was a positive experience. This is because we learn through challenge and difficulty. That's **(11)** ..... we get our wisdom. Few of **(12)** ..... , unfortunately, understand **(13)** ..... fact. For example, many people believe they suffer from stress at work, and take time off **(14)** ..... a result. Yet it has been found in some companies that by far **(15)** ..... healthiest people are those with the most responsibility. So next time you're in a stressful situation, just remember that it will be a positive learning experience and could also benefit your health!

References

Abney, S. P., & Johnson, M. (1991). Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research*, *20*, 233–250.

Alemán Bañón, J., Miller, D., & Rothman, J. (2017). Morphological variability in second language learners: An examination of electrophysiological and production data. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(10), 1509–1536.

Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(3), 714–735.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Lawrence Erlbaum Associates Publishers.

Armstrong, A., Bulkes, N. Z., & Tanner, D. (2018). Quantificational cues modulate the processing of English subject-verb agreement by native Chinese speakers: An ERP study. *Studies in Second Language Acquisition, 40*, 731–754.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The immediate use of gender information: eyetracking evidence of the time-course of pronoun resolution. *Cognition, 76*, 13–26.

Ashby, J., Treiman, R., Kessler, B., & Rayner, K. (2006). Vowel processing during silent reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 416–424.

Augurzky, P. (2006). *Attaching Relative Clauses in German: The Role of Implicit and Explicit Prosody in Sentence Processing.* Leipzig: MPI series in human cognitive and brain sciences, 77.

August, G. (2006). So, what's behind adult English second language reading? *Bilingual Research Journal, 30*, 245–264.

Autry, K. S., & Levine, W. H. (2014). A fan effect in anaphor processing: Effects of multiple distractors. *Frontiers in Psychology, 5*, Article 818.

Badecker, W., & Lewis, R. L. (2007). A new theory and computational model of working memory in sentence production: Agreement errors as failures of cue-

based retrieval. University of California, San Diego: Talk at the 20th CUNY Conference on Human Sentence Processing.

Badecker, W., & Lewis, R. L. (2007). A new theory and computational model of working memory in sentence production: agreement errors as failures of cue-based retrievals. *Proceedings of the CUNY Sentence Processing Conference*, San Diego, CA.

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 748–769.

Bader, M. (1998). "Prosodic influences on reading syntactically ambiguous sentences," in *Reanalysis in Sentence Processing*, eds J. Fodor and F. Ferreira (Dordrecht: Springer), 1–46. doi: 10.1007/978-94-015-9070-9_1

Bader, M., Meng, M., & Bayer, J. (2000). Case and reanalysis. *Journal of Psycholinguistic Research*, *29*(1), 37–52.

Baker, C. L. (1995). *English syntax*. 2nd ed. Cambridge, MA: MIT Press.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science, 7*(3), 136–141.

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, *17*, 137–153.

Bates, E. (1999). Plasticity, localization and language development. In S. H. Broman & J. M. Fletcher (Eds.), *The changing nervous system: Neurobehavioral consequences of early brain disorders* (pp. 214–253). New York: Oxford University Press.

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. Journal of Memory and Language, 30(6), 644-663.

Beach, C. M., Katz, W. F., & Skowronski, A. (1996). Children's processing of prosodic cues for phrasal interpretation. *Journal of the Acoustical Society of America*, *99*(2), 1148–1160.

Bel, A., & García-Alcaraz, E. (2015). Subject pronouns in the L2 Spanish of Moroccan Arabic speakers. In T. Judy & S. Perpiñán (Eds.), *The Acquisition of Spanish in Understudied Language Pairings* (pp. 201–232). Amsterdam: John Benjamins.

Bel, A., García-Alcaraz, E., & Rosado, E. (2016a). Reference comprehension and production in bilingual Spanish: The view from null subject languages. In A. A. de la Fuente, E. Valenzuela, & C. Martínez Sanz (Eds.), *Language Acquisition Beyond Parameters* (pp. 37–70). Amsterdam: John Benjamins.

Bel, A., Sagarra, N., Comínguez, J. P., & García-Alcaraz, E. (2016b). Transfer and proficiency effects in L2 processing of subject anaphora. *Lingua, 184*, 134–159.

Benedetto, S., Carbone, A., Pedrotti, M., Le Fevre, K., Bey, L. A. Y., & Baccino, T. (2015). Rapid serial visual presentation in reading: The case of Spritz. *Computers in Human Behavior, 45*, 352–358. https://doi.org/10.1016/j.chb.2014.12.043

Bernard, M. L., Chaparro, B. S., & Russell, M. (2001). Examining automatic text presentation for small screens. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 45*(6), 637–639.

Bernhardt, E., Kamil, M. (1995). Interpreting relationships between L1 and L2 reading: Consolidating the linguistic threshold and the linguistic interdependence hypotheses. *Applied Linguistics, 16*, 15–34.

Biau, E., Fromont, L. A., & Soto-Faraco, S. (2018). Beat gestures and syntactic parsing: An ERP study. *Language Learning, 68*, 102–126.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.

Blackwell, A., Bates, E., & Fisher, D. (1996), The time course of grammaticality judgment. *Language and Cognitive Processes, 11*(4), 337–406.

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology, 23*(1), 45–93.

Boland, J., Acker, M., & Wagner, L. (1998). The use of gender features in the resolution of pronominal anaphora. *Cognitive Science Technical Report, 17*, The Ohio State University Center for Cognitive Science.

Bock, J. K., & Eberhard, K. M. (1993). Meaning, sound and syntax in English number agreement. *Language and Cognitive Processes, 8*(1), 57–99.

Bock J. K., Nicol, J., & Cutting, J. C. (1999). The ties that bind: Creating number agreement in speech. *Journal of Memory and Language, 40*, 330–346.

Bock J. K., Eberhard, K. M., & Cutting, J. C. (2004). Producing number agreement: How pronouns equal verbs. *Journal of Memory and Language, 51*, 251–278.

Bock, J. K., Butterfield, S., Cutler, A., Cutting, J., Eberhard K. M., & Humphreys, K. R. (2006). Number agreement in British and American English: Disagreeing to agree collectively. *Language, 82*, 64–113.

Booth, P., Clenton, J., & Van Herwegen, J. (2018). Semantic and syntactic judgements: investigating the extent to which L1 learner choices influence the L2. *Systems, 78*, 54–64.

Bradley, D., Fernández, E. M., and Taylor, D. (2003). "Prosodic weight versus information load in the relative clause attachment ambiguity," in *Paper Presented at the 16th CUNY Conference on Human Sentence Processing*, (Cambridge, MA).

Branco, A., McEnery, T., & Mitkov, R. (2005). *Anaphora processing: Linguistic, cognitive, and computational modelling*. Amsterdam, Philadelphia: John Benjamins.

Brysbaert, M., & Mitchell, D. C. (1996). Modifier attachment in sentence parsing: Evidence from Dutch. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 49A*(3), 664–695.

Canal, P., Garnham, A., & Oakhill, J. (2015). Beyond gender stereotypes in language comprehension: self sex-role descriptions affect the brain's potentials associated with agreement processing. *Frontiers in Psychology, 6,* 1664–1078.

Caramazza, A., Grober, E., Garvey, C., & Yates, J. (1977). Comprehension of anaphoric pronouns. *Journal of Verbal Learning & Verbal Behavior, 16*(5), 601–609.

Carreiras, M., & Clifton, C. (1993). Relative clause interpretation preferences in Spanish and English. *Language & Speech, 36*, 353–372.

Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49A*(3), 639–663.

Carreiras, M., & Clifton, C. J. (1999). Another word on parsing relative clauses: Eye-tracking evidence from Spanish and English. *Journal of Memory and Cognition, 7,* 826–833.

Carrell, P. (1991). Second language reading: Reading ability or language proficiency? *Applied Linguistics, 12,* 159–179.

Cattell, J. M. (1947). Experiments on the association of ideas. In A. T. Poffenberger (Ed.), *James McKeen Cattell, man of science: Vol. I. Psychological research* (pp. 95-102). Lancaster, PA: Science Press. (Original work published 1887)

Chanales, A. J. H., Dudukovic, N. M., Richter, F. R., & Kuhl, B. A. (2019) Interference between overlapping memories is predicted by neural states during learning. *Nat Commun* 10**,** 5363 (2019) doi:10.1038/s41467-019-13377-x

Charnavel, I., & Sportiche, D. (2016). Anaphor binding: What French inanimate anaphors show. *Linguistic Inquiry 47*(1), 35–87.

Chen, L., Shu, H., Liu, Y., & Zhao, J. (2007). ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition, 10,* 161–174.

Cho, M. (2018). Task complexity, modality, and working memory in L2 task performance. *System, 72,* 85–98.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

Chomsky, N., & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology, Volume 2* (pp. 269–321). New York: Wiley.

Choy, J., & Thompson, C. K. (2010). How lexical processing deficits affect sentence comprehension in agrammatic Broca's aphasia. *Procedia - Social and Behavioral Sciences, 6,* 198–199.

Chow, W.-Y., Lewis, S., & Phillips, C. (2014). Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology, 5,* Article 630.

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *Quarterly Journal of Experimental Psychology, 69*(5), 817–828.

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences, 4*(5), 170–178.

Cirilo, R. K. (1981). Referential coherence and text structure in story comprehension. *Journal of Verbal Learning & Verbal Behavior*, *20*(3), 358–367.

Clark, M. (1979). Reading in Spanish and English: Evidence from adult ESL students. *Language Learning*, *29*, 121–150.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27*(1), 3–42.

Clifton, C., Frazier, L., & Deevy, P. (1999). Feature manipulation in sentence comprehension. *Rivista Di Linguistica*, *11*(1), 11–39.

Clifton, C., Carlson, K., & Frazier, L. (2002). Informative prosodic boundaries. *Language & Speech*, *45*, 87–114.

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *Child Development, 61*, 1584–1595.

Cowan, J. R. (1980). The significance of parallel function in the assignment of intrasentential anaphora. In J. Kreiman & A. E. Ojeda (Eds.), *Papers from the parasession on pronouns and anaphora* (pp. 110–124). Chicago: Chicago Linguistic Society.

Cromer, W. (1970). The difference model: A new explanation for some reading difficulties. *Journal of Educational Psychology*, *61*(6), 471–483.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*(2), 222–251.

Cummins, J. (1991). Interdependence of first- and second-language proficiency in bilingual children. In E. Bialystok (Ed.), *Language processing in bilingual children*. Cambridge: Cambridge University Press.

Cunnings, I., & Felser, C. (2013). The role of working memory in the processing of reflexives. *Language and Cognitive Processes, 28*, 188–219.

Cunnings, I., & Sturt, P. (2014). Coargumenthood and the processing of reflexives. *Journal of Memory and Language*, *75*, 117–139.

Cunnings, I., Fotiadou, G., & Tsimpli, I. (2017). Anaphora resolution and reanalysis during L2 sentence processing: evidence from the visual world paradigm. *Studies in Second Language Acquisition, 39*(4), 621–652.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40*(2), 141–201.

DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed), *Cognition and second language instruction* (pp. 125–151). New York: Cambridge University Press.

Dekydtspotter, L., Schwartz, B., & Sprouse, R. (2006). The Comparative fallacy in L2 processing research. In the Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006): The Banff Conference.

Dekydtspotter, L., Donaldson, B., Edmonds, A. C., Fultz, A. L., & Petrush, R. A. (2008). Syntactic and prosodic computations in the resolution of relative clause attachment ambiguity by English-French learners. *Studies in Second Language Acquisition, 30*(4), 453–480.

Dillon B. (2011). *Structured Access in Sentence Comprehension*. Unpublished Ph.D. dissertation, University of Maryland.

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C., (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, *69(2)*, 85–103.

Dillon B. (2014). Syntactic memory in the comprehension of reflexive dependencies: an overview. *Language and Linguistics Compass*, *8*, 171–187.

Dillon, B., Chow, W. Y., & Xiang, M. (2016). The Relationship Between Anaphor Features and Antecedent Retrieval: Comparing Mandarin Ziji and Ta-Ziji. *Frontiers in Psychology*, *6*, Article 1966.

Dinçtopal-Deniz, N. (2010). Relative clause attachment preferences of Turkish L2 speakers of English: Shallow parsing in the L2? In B. VanPatten & J. Jegerski (Eds.), *Research on second language processing and parsing* (pp. 27–63). Amsterdam: John Benjamins.

Domahs, U., Wiese, R., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). The processing of German word stress: evidence for the prosodic hierarchy. *Phonology, 25*(1), 1-36.

Dowhower, S. L. (1987). Effects of repeated reading on second-grade transitional readers' fluency and comprehension. *Reading Reading Quarterly*, *22*(4), 389–406.

Dowhower, S. L. (1991). Speaking of prosody: Fluency's unattended bedfellow. *Theory Into Practice, 30*(3), 165–175.

Drenhaus, H., Saddy, D., & Frisch, S. (2005). Processing negative polarity items: When negation comes through the backdoor. In S. Kepser & M. Reis (Eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives* (pp. 145–165). Berlin: de Gruyter.

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Reading Research Quarterly, 38*(1), 78–103.

Duffy, S. A., & Keir, J. A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition, 32*(4), 551–559.

Dussias, P. E. (2001). Sentence Parsing in Fluent Spanish-English Bilinguals. In J. L. Nicol (Ed.), *Explaining linguistics. One mind, two languages: Bilingual language processing* (pp. 159–176). Blackwell Publishing.

Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners: Some effects of bilinguality on LI and L2 processing strategies. *Studies in Second Language Acquisition, 25*(4), 529–557.

Dussias, P. E., Cramer Scaltz, T. R. (2008). Spanish-English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta Psychologica, 128*, 501–513.

Dwivedi, V. D. (2013). Interpreting quantifier scope ambiguity: evidence of heuristic first, algorithmic second processing. *PLoS ONE, 8*(11): e81461.

Ehrlich, K. (1980). Comprehension of pronouns. The Quarterly Journal of Experimental Psychology, 32(2), 247–255.

Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of Verbal Learning & Verbal Behavior, 22*(1), 75–87.

Ehrlich, K., Fernández, E., Fodor, J. D., Stenshoel, E., Vinereanu, M. (1999). Low attachment of relative clauses: New data from Swedish, Norwegian and Romanian. Paper presented at the 12th CUNY Conference on Human Sentence Processing, New York, NY.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception and Psychophysics, 67*, 224–238.

Engelmann, F., Jäger, L., and Vasishth, S. (2015). "Cue confusion and distractor prominence explain inconsistent effects of retrieval interference in human sentence processing," in *Proceedings of the 13th International Conference on Cognitive Modeling* (Groningen), 192–193.

Engelmann, F., Jäger, L.A. and Vasishth, S. (2019), The Effect of Prominence and Cue Association on Retrieval Processes: A Computational Account. *Cogn Sci,* 43: e12800. https://doi.org/10.1111/cogs.12800.

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language, 54*(4), 541–553.

Felser, C. & Cunnings, I. (2012). Processing reflexives in a second language: The timing of structural and discourse-level constraints. *Applied Psycholinguistics, 33*, 571–603.

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics, 24*(3), 453–489.

Felser, C., Sato F. M., & Bertenshaw, N. (2009). The on-line application of Binding Principle A in English as a second language. *Bilingualism, 12*, 485–502.

Fernández, E. M. (2007). How might a rapid serial visual presentation of text affect the prosody projected implicitly during silent reading? In *Conferências do V Congresso Internacional da Associaçao Brasiliera de Lingüistica* (Vol. 5, pp. 117–154).

Fernández, E. M., Bradley, D., Igoa, J. M., and Teira, C. (2003). "Prosodic phrasing patterns in the RC-attachment ambiguity: effects of language, RC-length, and position," in *Paper Presented at the AMLaP*-2003 Conference, (Glasgow).

Fernández, E. M., & Sekerina, I. (2015). The interplay of visual and prosodic information in the attachment preferences of semantically shallow relative clauses. In Frazier L., Gibson E. (Eds.), *Explicit and Implicit Prosody in Sentence processing* (pp. 241–261), Cham: Springer International Publishing.

Ferreira, F. (1988). Planning and timing in sentence production: The syntax-to-phonology conversion. *Dissertation Abstracts International, 49*(5-B), 1971.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47,* 164–203.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11,* 11–15.

Ferreira, F., & Cokal, D. (2015). Sentence processing. In G. Hickok and S. Small (Eds.), *Neurobiology of Language*: Elsevier.

Ferreira, F., & Patson, N. D. (2007). The "good enough" approach to language comprehension. *Language and Linguistics Compass, 1*(1-2), 71–83.

Fific, M., Nosofsky, R. M., & Townsend, J. T. (2008). Information-processing architectures in multidimensional classification: A validation test of the systems factorial technology. *Journal of Experimental Psychology: Human Perception and Performance, 34*(2), 356–375. https://doi.org/10.1037/0096-1523.34.2.356

Flege, J.E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustic Society of America, 76*(3), 708–721.

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research, 27*(2), 285–319.

Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. *Proceedings of the 1st International Conference on Speech Prosody* (pp. 83–88).

Foote, R. (2011). Integrated knowledge of agreement in early and late English-Spanish bilinguals. *Applied Psycholinguistics, 32,* 187–220.

Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language, 56*(3), 357–383.

Franceschina, F. (2001a). Against an L2 morphological deficit as an explanation for the differences between native and non-native grammars. In S. H. Foster-Cohen, & A. Nizegorodcew (Eds.), *EUROSLA yearbook*, 1 (pp. 143–158). Amsterdam: Benjamins.

Franceschina, F. (2001b). Morphological or syntactic deficits in near–native speakers? An assessment of some current proposals. *Second Language Research*, *17*, 213–247.

Franceschina, F. (2002). Case and phi-feature agreeement in advanced L2 Spanish grammars. In S. H. Foster-Cohen, R. Ruthenberg, & M. L. Poschen (Eds.), *EUROSLA yearbook*, 2 (pp. 71–86). Amsterdam: Benjamins.

Franceschina, F. (2005). *Fossilized second language grammars: The acquisition of syntactic gender*. Amsterdam: Benjamins.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. Bloomington, Indiana: Indiana University Linguistics Club.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: a new two-stage parsing model. *Cognition, 6,* 291–325.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178–210.

Frazier, L., & Clifton, C., Jr. (1996). *Construal*. The MIT Press.

Frazier, L., Carlson, K., & Clifton, C., Jr. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences, 10*(6), 244–249.

Frenck-Mestre, C. (2002). An on-line look at sentence processing in the second language. In R. Heredia & J. Altarriba (eds.), *Bilingual Sentence Processing*, 218–236. Amsterdam: Elsevier.

Frenck-Mestre, C. & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *The Quarterly Journal of Experimental Psychology, 50A*, 119–148.

Fukkink, R. G., Hulstijn, J., & Simis, A. (2005). Does Training in Second-Language Word Recognition Skills Affect Reading Comprehension? An Experimental Study. *Modern Language Journal, 89*(1), 54–75.

Fultz, A. L. (2007). Prosody in syntactic ambiguation in English-French interlanguage. In the Proceedings of the Annual Boston University Conference on Language Development, pp. 394–405.

Fultz A. L. (2009). *Prosody in lexical and syntactic disambiguation in English-French interlanguage*, Ph.D. thesis, Indiana University, Bloomington, IN.

Galecki, A., & Burzykowski, T. (2013). *Linear mixed effects models using R: A step by step approach*. New York: Springer.

Gardelle, L. (2012). Anaphora, anaphor and antecedent in nominal anaphora: definitions and theoretical implication. *Cercles*, *22*, 25–40.

Garnham, A. (2001). *Essays in cognitive psychology. Mental models and the interpretation of anaphora.* Psychology Press.

Garnham, A., Oakhill, J., & Cruttenden, H. (1992). The role of implicit causality and gender cue in the interpretation of pronouns. *Language and Cognitive Processes, 7*(3-4), 231–255.

Garnham, A., Traxler, M., Oakhill, J., & Gernsbacher, M. A. (1996). The locus of implicit causality effects in comprehension. *Journal of Memory and Language, 35*(4), 517–543.

Garrod, S., & Sanford, A. J. (1985). On the real-time character of interpretation during reading. *Language and Cognitive Processes*, *1*(1), 43–59.

Garrod, S. C., & Sanford, A. J. (1994). Resolving sentences in a discourse context: How discourse representation affects language understanding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 675–698). Academic Press.

Garrod, S., Freudenthal, D., & Boyle, E. A. (1994). The role of different types of anaphor in the on-line resolution of sentences in a discourse. *Journal of Memory and Language, 33*(1), 39–68.

Garrod, S., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. *Journal of Memory and Language, 42*(4), 526–544.

Gee, J., & Grosjean, F. (*1983*). Performance structures: a psycholinguistic and linguistic appraisal. *Cognitive Psychology, 15,* 411–458.

Gerber-Morón, O., Szarkowska, A., & Woll, B. (2018). The impact of text segmentation on subtitle reading. *Journal of Eye Movement Research*, *11*(4), 1–18.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. Cognition, *32*, 99–156.

Ghilzai, S. (2017). Sensitivity to morphosyntactic features in L2 sentence processing: Evidence from Persian, Urdu, and Japanese. *Ilsc*, *2*, 86-103.

Gibson, E., Pearlmutter, N., Canseco-Gonzalez, E., & Hickok, G. (1996). Recency preference in the human sentence processing mechanism. *Cognition, 59*(1), 23–59.

Gibson, E. (2000). The dependency locality theory: a distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium* (pp. 95–126). The MIT Press.

Gibson, E., Desmet, T., Grodner, D., Watson, D., & Ko, K. (2005). Reading relative clauses in English. *Cognitive Linguistics, 16*(2), 313–353.

Gilboy, E., Sopena, J.-M., Clifton, C., & Frazier, L. (1995). Argument structure and association preferences in Spanish and English complex NPs. *Cognition, 54*(2), 131–167. https://doi.org/10.1016/0010-0277(94)00636-Y

Gleitman, L. R., & Wanner, E. (1982). Language acquisition: The state of the state of the art. In E. Wanner, & L. R. Gleitman (Eds.), *Language acquisition: State of the art* (pp. 3–48). New York: Cambridge University Press.

Goh, W. D., & Lu, S. H. X. (2012). Testing the myth of the encoding–retrieval match. *Memory & Cognition*, *40*(1), 28–39. https://doi.org/10.3758/s13421-011-0133-9

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science, 17*(3), 311–347.

Goss, S., & Nakayama, M. (2011). Prosody and comprehension in oral reading by L2 Japanese learners. *Second Language*, *10*, 33–50.

Grabe, W., & Stroller, F. L. (2002). *Teaching and researching reading*. London: Longman.

Greene, S. B., McKoon, G., & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(2), 266–283.

Grillo, N., Aguilar, M., Roberts, L., Santi, A., & Turco, G. (2018). Prosody of classic garden path sentences: The horse raced faster when embedded. Paper presented at Speech Prosody 2018, Poznan, Poland.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive Science, 29*(2), 261–290.

Grodner, D., Gibson, E., & Watson, D. (2005). The influence of contextual contrast on syntactic processing: Evidence for strong-interaction in sentence comprehension. *Cognition, 95*(3), 275–296.

Groen, M. A., Veenendaal, N. J., & Verhoeven, L. (2019). The role of prosody in reading comprehension: evidence from poor comprehenders. *Journal of Research in Reading, 42*, 37– 57.

Grosz, B. J., Aravind K. J., & Scott W. (1983). Providing a unified account of definite noun phrases in discourse. In *Proc. 21st Annual Meeting of the ACL* (pp. 44–50), Association of Computational Linguistics.

Grosz, B. J., Aravind K. J., & Scott W. (1986). *Towards a computational theory of discourse interpretation*. Unpublished Manuscript.

Grosz, B.J., Sidner, C.L. (1986). Attention, intention, and the structure of discourse. *Computational Linguistics, 12*, 175–204.

Gundel, J. K. (1999a). On different kinds of focus. In Bosch, & van der Sandt (Eds.), *Focus: Linguistic, Cognitive, and Computational Perspectives* (pp. 293–305). Cambridge: Cambridge University Press.

Gundel, J. K. (1999b). Topic, focus and the grammar pragmatics interface. In J. Alexander, N. Han, & M. Minnick (Eds.), *Proceedings of the 23rd Annual Penn Linguistics Colloquium, Vol. 6.1.*, (pp. 185–200). Penn Working Papers in Linguistics.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language, 69*, 274–307.

Guo, J., Taomei, G., Yan, Y., Nan, J., Danling, P., (2009). ERP evidence for different strategies employed by native speakers and L2 learners in sentence processing. *Journal of Neurolinguistics, 22,* 123–134.

Haegeman, L. (1991). *Introduction to government and binding theory*. Oxford & Cambridge, MA: Basil Blackwell.

Hagoort, P., Brown, C. M., & Groothsen, J., (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8,* 439–483.

Hagoort, P., Brown, C. M., & Osterhout, L., (1999). The neurocognition of syntactic processing. In C.M. Brown, & P. Hagoort (Eds.), *The Neurocognition of Language* (pp. 273–316). Oxford: Oxford University Press.

Harris, J., Jun, S.-A., and Royer, A. (2016). Implicit prosody pulls its weight: recovery from garden path sentences. *Speech Prosody* 2016, 207–211. doi: 10.21437/SpeechProsody.2016-43

Haskell, T. R., & MacDonald, M. C. (2003). Conflicting cues and competition in subject-verb agreement. *Journal of Memory and Language, 48*(4), 760–778.

Hawkins, R. & Franceschina, F. (2004). Explaining the acquisition and non-acquisition of determiner-noun gender concord in French and Spanish. In P. Prévost & J. Paradis (Eds.), *The acquisition of French in different contexts*. Amsterdam/Philadelphia: John Benjamins.

Hayes, B. (1989). The prosodic hierarchy in meter. In P. Kiparsky & G. Youmans (Eds.), *Rhythm and meter* (pp. 201–260). New York: Academic Press.

Heffner, C. C., & Slevc, L. R. (2015). Prosodic structure as a parallel to musical structure. *Frontiers in Psychology, 6,* Article 1962.

Hellbernd, N., & Sammler, D. (2016). Prosody conveys speaker's intentions: Acoustic cues for speech act perception. *Journal of Memory and Language, 88,* 70–86.

Hemforth, B., Konieczny, L., & Scheepers, C. (2000a). Modifier attachment: relative clauses and coordinations. In B. Hemforth, & L. Konieczny (Eds.) *German sentence processing,* (pp. 159–163), Dordrecht, NL: Kluwer.

Hemforth, B., Konieczny, L., Seelig, H., & Walter, M. (2000b). Case matching and relative clause attachment. *Journal of Psycholinguistic Research, 29/1*, 81–88.

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. Nature reviews. *Neuroscience, 11,* 523–532.

Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General, 124*(1), 62–82.

Hijikata, Y. (2005). The chunking process and reading comprehension of Japanese EFL learners. *Annual Review of English Language Education in Japan, 16,* 61–70.

Hirose, Y. (2003). Recycling prosodic boundaries. *J. Psycholinguist Res.* 32, 167–195. doi: 10.1023/A:1022448308035

Hoffmann, C. (2012). *Cohesive profiling. Meaning and interaction in personal weblogs*. Amsterdam – Philadelphia: Benjamins.

Hoover, M. L., & Dwivedi, V. D. (1998). Syntactic processing in skilled bilinguals. *Language Learning,* 48, 1–29.

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research, 22*(3), 369–397.

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching, 53,* 277–306.

Hopp, H. (2017). The processing of English which-questions in adult L2 learners: Effects of L1 transfer and proficiency. *Zeitschrift fur Sprachwissenschaft, 36,* 107–134.

Hoyte, K. J., Brownell, H., & Wingfield, A. (2009). Components of speech prosody and their use in detection of syntactic structure by older adults. *Experimental Aging Research, 35*(1), 129–151.

Hudson-D'Zmura, S., & Tanenhaus, M. K. (1998). Assigning antecedents to ambiguous pronouns: The role of the center of attention as the default assignment. In M. Walker, A. K. Joshi, & E. Prince (Eds.), *Centering Theory in Discourse* (pp. 199–226). Oxford: Oxford University Press.

Huang, Y. (2000). *Anaphora: A cross-linguistic approach*. Oxford: Oxford University Press.

Hwang, H., and Schafer, A. J. (2009). Constituent length affects prosody and processing for a dative NP ambiguity in Korean. *J. Psycholinguist Res.* 38, 151–175. doi: 10.1007/s10936-008-9091-1

Inkelas, S. (1990). Prosodic Constituency in the Lexicon. New York, NY: Routledge.

Jackson, C. (2008). Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence processing. *Language Learning*, 58 (4), 875– 909.

Jackson, C. N., & O'Brien, M. (2011). The interaction between prosody and meaning in second language speech production. *Teaching German, 44*(1), 1–11.

Jackson, C. N., & van Hell, J. G. (2011). The effects of L2 proficiency level on the processing of wh-questions among Dutch second language speakers of English. *IRAL, International review of applied linguistics in language teaching: Revue internationale de linguistique appliquee enseignement des langues. Internationale Zeitschrift fur angewandte Linguistik in der Spracherziehung*, *49*(3), 195–219.

Jacob, G., & Felser, C. (2016). Reanalysis and semantic persistence in native and non-native garden-path recovery. *The Quarterly Journal of Experimental Psychology, 69*, 907–925.

Jäger, L. A., Benz, L., Roeser, J., Dillon, B. W., & Vasishth, S. (2015). Teasing apart retrieval and encoding interference in the processing of anaphors. *Frontiers in Psychology, 6,* Article 506.

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, 94, 305– 315.

Jegerski, J. (2014). Self-paced reading. *Research Methods in Second Language Psycholinguistics,* 20–49: Taylor and Francis.

Jeong, H., Sugiura, M., Sassa, Y., Haji, T., Usui, N., Taira, M., Horie, K., Sato, S., & Kawashima, R. (2007). Effect of syntactic similarity on cortical activation during second language processing: A comparison of English and Japanese among native Korean trilinguals. *Human Brain Mapping, 28*, 194–204.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics, 25*, 603–634.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning, 57*, 1–33.

Jiang, N. (2018). *Second language acquisition: an introduction*. New York, NY: Routledge.

Jiang, X. (2011). The role of first language literacy and second language proficiency in second language reading comprehension. *The Reading Matrix, 11*.

Joh, J., & Plakans, L. (2017). Working memory in L2 reading comprehension: The influence of prior knowledge. *System, 70*, 107–120.

Johns, C. L., Matsuki, K., & Van Dyke, J. A. (2015). Poor readers' retrieval mechanism: Efficient access is not dependent on reading skill. *Frontiers in Psychology, 6*, Article 1552.

Jun, S.-A., and Bishop, J. (2015). Priming implicit prosody: prosodic boundaries and individual differences. *Lang. Speech.* 58, 459–473. doi: 10.1177/0023830914563368

Jung, J. (2018). Effects of task complexity and working memory capacity on L2 reading comprehension. *System, 74*, 21–37.

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L. J., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology, 24*(2), 252–293.

Jusczyk, P. W. (2003). Chunking language input to find patterns. In D. H. Rakison & L. M. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (p. 27–49). Oxford University Press.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122–149.

Kadota, S. (1982). Some psycholinguistic experiments on the process of reading comprehension. *Journal of Assumption Junior College, 9*, 49–70.

Kadota, S., & Tada, M. (1992). Reading comprehension and processing units in English and Japanese. *Annual Bulletin of Research Institute for Social Science, 22*, 137–153.

Kadota, S., Yoshida, S., & Yoshida, H. (1999). Processing units in EFL reading: An effect of presentation units on comprehension rate and time. *Annual Review of English Language Education in Japan, 10,* 61–71.

Kaiser, E., Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition, 112*(1), 55–80.

Kalbertodt, J., Primus, B., & Schumacher, P. B. (2015). Punctuation, prosody, and discourse: afterthought vs. right dislocation. *Frontiers in psychology, 6,* Article 1803.

Kamide, Y., & Mitchell, D. (1997). Relative clause attachment: nondeterminism in Japanese parsing. *Journal of Psycholinguistic Research, 26,* 247–254.

Kang, T. J., & Muter, P. (1989). Reading dynamically displayed text. *Behavior & Information Technology, 8*(1), 33–42.

Kang, O., Johnson, D. O., Kermad, A. (2021). *Second language prosody and computer modeling*. Routledge.

Katz, W. F., Beach, C. M., Jenouri, K., & Verma, S. (1996). Duration and fundamental frequency correlates of phrase boundaries in productions by children and adults. *The Journal of the Acoustical Society of America, 99*(5), 3179–3191.

Keating, G. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish. *Language Learning*, *59*, 503–535.

Kilborn, K., & Cooreman, A. (1987). Sentence interpretation strategies in adult Dutch–English bilinguals. *Applied Psycholinguistics*, *8*(4), 415–431.

Kim, J. H., & Christianson, K. (2017). Working memory effects on L1 and L2 processing of ambiguous relative clauses by Korean L2 learners of English. *Second Language Research*, *33*(3), 365–388.

Kim, S. Y., Qi, T., Feng, X., Ding, G., Liu, L., & Cao, F. (2016). How does language distance between L1 and L2 affect the L2 brain network? An fMRI study of Korean–Chinese–English trilinguals. *Neuro Image, 129,* 25–39.

Kim, S. Y., Liu, L., & Cao, F. (2017). How does first language (L1) influence second language (L2) reading in the brain? Evidence from Korean-English and Chinese-English bilinguals. *Brain and Language, 171,* 1–13.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language, 40*(2), 153–194.

Klassen, J., & Wagner, M. (2017). Prosodic prominence shifts are anaphoric. *Journal of Memory and Language, 92*, 305–326.

Klin, C. M., Guzmán, A. E., Weingartner, K. M., & Ralano, A. S. (2006). When anaphor resolution fails: Partial encoding of anaphoric inferences. *Journal of Memory and Language, 54*(1), 131–143.

Kocaarslan, M. (2019). The effects of reading rate, accuracy and prosody on second grade students' oral retelling. *Acta Psychologica, 197*, 86–93.

Koda, K. (1989). The effects of transferred vocabulary knowledge on the development of L2 reading proficiency. *Foreign Language Annals, 22*, 529–540.

Koriat, A., Greenberg, S. N., & Kreiner, H. (2002). The extraction of structure during reading: Evidence from reading prosody. *Memory & Cognition, 30*(2), 270–280.

Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology, 50*(2), 194–231.

Kreiner, H. (2006). The Role of Reading Prosody in Syntactic and Semantic Integration: Evidence from Eye Movements. Paper presented at the Discourse–Prosody Interface (IDP) Conference, Aix-en-Provence, France

Kreiner, H., Sturt, P., & Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language, 58*(2), 239–261.

Kreiner, H., Garrod, S., & Sturt, P. (2013). Number agreement in sentence comprehension: The relationship between grammatical and conceptual factors. *Language and Cognitive Processes, 28,* 829–874.

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology, 95*(1), 3–21.

Kush, D., Lidz, J., & Phillips, C. (2015). Relation-sensitive retrieval: Evidence from bound variable pronouns. *Journal of Memory and Language, 82,* 18–40.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293–323.

Ladd, D. R. (1984). English compound stress. In D. Gibbon and H. Richter (Eds.), *Intonation, Accent, and Rhythm* (pp. 253–266). Berlin: de Gruyter.

Ladd, D. (2008). *Intonational phonology (Cambridge studies in linguistics).* Cambridge: Cambridge University Press

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa, 7*, 107–122.

LeVasseur, V. M., Macaruso, P., Palumbo, L. C., & Shankweiler, D. (2006). Syntactically cued text facilitates oral reading fluency in developing readers. *Applied Psycholinguistics*, *27*, 423–445.

LeVasseur, V. M., Macaruso, P., & Shankweiler, D. (2008). Promoting gains in reading fluency: A comparison of three approaches. *Reading and Writing, 21*, 205–230.

Lewis, R. L., & Vasishth, S. (2005). An Activation-Based Model of Sentence Processing as Skilled Memory Retrieval. *Cognitive Science, 29*(3), 375–419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences, 10*(10), 447–454.

Liang, L., Wen, Y., & Dong, Y. (2018). Gender constraint in L1 and L2 reflexive pronoun resolution by Chinese-English bilinguals. *Journal of Neurolinguistics*, *45*.

Lieberman, P. (1967). Intonation, perception, and language. *M.I.T. Research Monograph, 38*, xiii, 210.

Lim, J. H., & Christianson, K. (2013). Integrating meaning and structure in L1–L2 and L2–L1 translations. *Second Language Research, 29*(3), 233–256. https://doi.org/10.1177/0267658312462019

MacWhinney, B. (1986). Toward a psycholinguistically plausible parser. In S. Thomason (Ed.), *Proceedings of the Eastern States Conference on Linguistics (ESCOL) 1986* (pp. 1–8). Columbus, OH: Ohio State University.

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language processing. *Studies in Second Language Acquisition, 27*, 53–78.

Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P., & Lee, C. S. (1992). Prosodic effects in minimal attachment. *The Quarterly Journal of Experimental Psychology*, *45*, 73–87.

Martin, A. E., & McElree, B. (2008). A content–addressable pointer mechanism underlies comprehension of verb–phrase ellipsis. *Journal of Memory and Language, 58*(3), 879–906.

Martin, A. E., & McElree, B. (2009). Memory operations that support language comprehension: Evidence from verb–phrase ellipsis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(5), 1231–1239.

Martin, A. E., & McElree, B. (2011). Direct–access retrieval during sentence comprehension: Evidence from sluicing. *Journal of Memory and Language, 64*(4), 327–343.

Martin A. E. (2018). Cue integration during sentence comprehension: Electrophysiological evidence from ellipsis. *PloS one, 13*(11), e0206616.

Matsui, T., Nakamura, T., Utsumi, A., Sasaki, A. T., Koike, T., Yoshida, Y., Harada, T., Tanabe, H. C., & Sadato, N. (2016). The role of prosody and context in sarcasm comprehension: Behavioral and fMRI evidence. *Neuropsychologia*, *87*, 74–84.

Maynell, L. (1999). Effect of pitch accent placement on resolving relative clause ambiguity in English. Paper presented at the 12th Annual CUNY Conference on Human Sentence Processing. New York, NY.

Maynell, L. (2000). Prosodic effects on relative clause attachment. Paper presented at the 13th Annual CUNY Conference on Human Sentence Processing. La Jolla, CA.

McCarthy, C. (2008). Morphological variability in the comprehension of agreement: An argument for representation over computation. *Second Language Research, 24*, 459–486.

McDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes, 9*, 157–201.

McDonald, J. L., & MacWhinney, B. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language, 34*(4), 543–566.

McElree, B., & Dosher, B. A. (1993). Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General, 122*(3), 291–315.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research, 29*(2), 111–123.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 817–835.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*(1), 67–91.

McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The psychology of learning and motivation: Vol. 46. The psychology of learning and motivation: Advances in research and theory* (pp. 155–200). Elsevier Academic Press.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. Psychological Review, *99*(3), 440–466.

Miller, G. & Isard, S. (1964). Free Recall of Self-Embedded English Sentences. *Information and Control*, *7*, 292-303.

Miller, J., & Schwanenflugel, P. J. (2006). Prosody of syntactically complex sentences in the oral reading of young children. *Journal of Educational Psychology*, *98*(4), 839–853.

Mitkov, R. (2002). *Anaphora Resolution*. London: Longman.

Molinaro, N., Su, J., & Carreiras, M. (2016). Stereotypes override grammar: Social knowledge in sentence comprehension. *Brain and Language*, *155*, 36–43.

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, *66*, 911–936

Nakassis, C., & Snedeker, J. (2002). Beyond sarcasm: Intonation and context as relational cues in children's recognition of irony. *Proceedings of the Twenty-Sixth Boston University Conference on Language Development,* 429–440. Cascadilla Press, Somerville, MA.

Nelson, C. A. (1999). Change and continuity in neurobehavioral development: Lessons from the study of neurobiology and neural plasticity. *Infant Behavior and Development*, 22, 415–429.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realizations in VO and OV languages. *Lingue e Linguaggio*, 2, 1–29.

Nguyen, A. T., Ingram, J., & Pensalfini, R. (2008). Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. *Journal of Phonetics*, *36*, 158–190.

Nickels, S., & Steinhauer, K. (2018). Prosody–syntax integration in a second language: Contrasting event-related potentials from German and Chinese learners of English using linear mixed effect models. *Second Language Research*, *34*(1), 9–37.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research, 18*(1), 5–19.

Nicol, J. L., Fodor, J. D., & Swinney, D. (1994). Using cross-modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(5), 1229–1238.

Nicol, J. L., Forster, K. I., & Veres, C. (1997). Subject-verb agreement processes in comprehension. *Journal of Memory and Language, 36*(4), 569–587.

Nicol, J. L., & Swinney, D. (2003). The psycholinguistics of anaphora. In A. Barss (Ed.), *Anaphora: A reference guide* (pp. 72-104). Maiden, MA: Blackwell.

O'Shea, L. J., & Sindelar, P. T. (1983). The effects of segmenting written discourse on the reading comprehension of low- and high-performance readers. *Reading Research Quarterly*, *18*(4), 458–465.

Oakhill, J., Garnham, A., & Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Memory and cognition*, *33*, 972–983.

Osterhout, L., Bersick, M. & Mckinnon, R. (1997). Brain potentials elicited by words: Word length and frequency predict the latency of an early negativity. *Biological psychology*, *46*, 143–168.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language, 34*(6), 739–773.

Osterhout, L., Kim, A., & Kuperberg, G. R. (2012). The neurobiology of sentence comprehension. In M. J. Spivey, K. McRae, & M. F. Joanisse (Eds.), *Cambridge handbooks in psychology. The Cambridge handbook of psycholinguistics* (pp. 365–389). Cambridge University Press.

Öztekin, I., Davachi, L., & McElree, B. (2010). Are representations in working memory distinct from representations in long-term memory? Neural evidence in support of a single store. *Psychological Science, 21*(8), 1123–1133.

Paige, D. D., Rasinski, T., Magpuri-Lavell, T., & Smith, G. S. (2014). Interpreting the relationships among prosody, automaticity, accuracy, and silent reading comprehension in secondary students. *Journal of Literacy Research*, *46*(2), 123–156.

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in LI and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition, 25*(4), 501–528.

Papadopoulou, D., & Tsimpli, I. M. (2005). Morphological cues in children's processing of ambiguous sentences: A study of subject/object ambiguities in Greek. In A. Brugos, M. R. Clark-Cotton & S. Ha (Eds.), BUCLD 29: Proceedings of the 29th annual Boston University Conference on Language Development (pp. 471–481). Boston: Cascadilla Press.

Park, G. P., & Chae, J. S. (2000). Testing the threshold hypothesis in foreign language reading. *Foreign Languages Education, 7*, 259–276.

Park, G. P. (2013). Relations among L1 reading, L2 knowledge, and L2 reading: revisiting the Threshold Hypothesis. *English Language Teaching, 6.*

Parker, D. (2019). Cue Combinatorics in Memory Retrieval for Anaphora. *Cognitive Science* 43 (3), 1–30.

Parker, D., Lago, S., & Phillips, C. (2015). Interference in the processing of adjunct control. *Frontiers in Psychology*, *6*, 1–13.

Parker, D., Shvartsman, M., & Van Dyke, J. A. (2015). *Agreement illusions are selective: Evidence from eye-tracking and modeling*. Paper presented at the Experimental Psycholinguistics Conference, Universidad Nacional de Educación a Distancia.

Parker D., & Phillips C. (2014). Selective priority for structure in memory retrieval. In *Poster at the 27th CUNY Conference on Human Sentence Processing*, The Ohio State University, Columbus, OH.

Parker, D., & Phillips, C. (2016). Negative polarity illusions and the format of hierarchical encodings in memory. *Cognition, 157,* 321–339.

Parker, D. & Phillips, C. (2017). Reflexive attraction in comprehension is selective. *Journal of Memory and Language, 94,* 272–290.

Parker, D., Shvartsman, M., & Van Dyke, J. A. (2017). The cue-based retrieval theory of sentence comprehension: New findings and new challenges. In L. Escobar, V. Torrens & T. Parodi (Eds.), *Language processing and disorders* (pp. 121–144). Newcastle: Cambridge Scholars Publishing.

Patil, U., Vasishth, S., & Lewis, R. L. (2016). Retrieval interference in syntactic processing: The case of reflexive binding in English. *Frontiers in Psychology, 7,* 1–18.

Patterson, C., Trompelt, H., & Felser, C. (2014). The online application of binding condition B in native and non-native pronoun resolution. *Frontier in Psychology, 5,* Article 147.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language, 41*(3), 427–456.

Perfetti, C.A. (1985) Reading Ability. Oxford University Press, New York.

Perfetti, C. A. (1988). Verbal efficiency in reading ability. In M. Daneman, G. E. Mackinnon, & T. G. Waller (Eds.), *Reading research: Advances in theory and practice, Vol. 6* (pp. 109–143). Academic Press.

Peterson, W. W. T. G., Birdsall, T., and Fox, W. (1954). The theory of signal detectability. *Trans. IRE Prof. Group Inf. Theory* 4, 171–212. doi: 10.1109/TIT.1954.1057460

Piaget, J. (1952). *The origins of intelligence in children* (M. Cook, Trans.). W W Norton & Co.

Piaget, J. (1977). *The Development of Thought. Equilibration of Cognitive Structures*. Oxford Basil Blackwell.

Pickering, M., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes, 6*(3), 229–259.

Pierrehumbert, J. B., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.),

Intentions in communication (pp. 271–311). Cambridge, MA: MIT Press.

Pollard, C., & Sag, I. A. (1992). Anaphors in English and the scope of binding theory. *Linguistic Inquiry, 23*, 261–303.

Phillips, C., Wagers, M. W., & Lau, E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. *Language*, *82*, 795–782.

Pratt, E. (2015). *Is cue-based retrieval "good enough"?: Agreement, comprehension, and implicit prosody in native and bilingual speakers of English* (Ph.D. dissertation). City University of New York.

Pratt, E., & Fernández, E. M. (2016). Implicit prosody and cue-based retrieval: L1 and L2 agreement and comprehension during reading. *Frontiers in Psychology, 7,* Article 1922.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 90*(6), 2956–2970.

Pulvermüller F., Shtyrov Y., Hasting A., & Carlyon R. P. (2008). Syntax as a reflex: Neurophysiological evidence for early automaticity of grammatical processing. *Brain and Language, 104*(1), 244–253.

Quirk, R., Sidney G., Geoffrey L., & Jan S. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.

Rasinski, T. V. (1989). Adult readers' sensitivity to phrase boundaries in texts. *Journal of Experimental Education, 58*(1), 29–40.

Rasinski, T. V. (1990). Effects of repeated effects and listening-while-reading on reading fluency. *The Journal of Educational Research*, *83*(3), 147–150.

Rasinski, T. V. (1999). Exploring a method for estimating independent, instructional, and frustration reading rates. *Reading Psychology*, *20*(1990), 61–69.

Rasinski, T. V., Padak, N., Linek, W., & Sturtevant, E. (1994). Effects of fluency development on urban second-grade readers. *The Journal of Educational Research*, *87*(3), 158–165.

Reinhart, T. (1976*). The syntactic domain of anaphora* (Ph.D. dissertation). MIT.

Reinhart, T. (1981). Definite NP anaphora and C-command domains. *Linguistic Inquiry, 12*, 605–635.

Reinhart, T., & Reuland, E. (1993). Reflexivity. *Linguistic Inquiry*, 657–720.

Reinhart, T. (2006). *Interface strategies – reference-set computation*. Cambridge, MA: The MIT Press.

Reuland, E. (2001b). Primitives of binding. *Linguistic Inquiry*, *32*, 439–492.

Reuland, E. (2011). *Anaphora and language design*. Cambridge, Massachusetts: MIT Press.

Reynolds, D., Garnham, A., & Oakhill, J. (2006). Evidence of immediate activation of gender information from a social role name. *Quarterly Journal of Experimental Psychology*, *59*, 886–903.

Roehm, D., Bornkessel, I., Haider, H., & Schlesewsky, M. (2005). When case meets agreement: Event-related potential effects for morphology-based conflict resolution in human language comprehension. *NeuroReport: For Rapid Communication of Neuroscience Research, 16*(8), 875–878.

Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics, 32*(2), 299–331.

Rodríguez, G. A. (2008). *Second Language Sentence Processing: Is It Fundamentally Different?* Unpublished Ph.D. dissertation, University of Pittsburgh, Pittsburgh.

Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2003). Assignment of reference to reflexives and pronouns in picture noun phrases: evidence from eye movements. *Cognition, 89*, B1-B13.

Runner, J. T., & Kaiser. E. (2005). Binding in Picture Noun Phrases: Implications for Binding Theory. In S. Müller (Ed), *The Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar - Department of Informatics, University of Lisbon* (pp. 594–613), Stanford: CSLI Publications.

Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2006). Processing reflexives and pronouns in picture noun phrases. *Cognitive Science, 30*, 193–241.

Sagarra, N., & Herschensohn, J. (2010). Proficiency and animacy effects on L2 gender agreement processes during comprehension. *Language Learning, 61*, 80–116.

Sagarra, N., & Herschensohn, J. (2013). Processing of gender and number agreement in late Spanish bilinguals. *International Journal of Bilingualism*, *17*(5), 607–627.

Samuels, S. J., Schermer, N., & Reinking, D. (1992). Reading fluency: Techniques for

making decoding automatic. In S. Jay Samuels & Alan E. Farstrup (Eds.), *What research has to say about reading instruction1* (2nd Edition, pp. 124–144). Newark, DE: International Reading Association.

Sanford, A. J., Garrod, S. C., Lucas, A., & Henderson, R. (1983). Pronouns without explicit antecedents. *Journal of Semantics, 2*, 303–318.

Sanford, A. J., Moar, K., & Garrod, S. C. (1988). Proper names as controllers of discourse focus. *Language and Speech, 31*(1), 43–56.

Scott, D. R. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America, 71*, 996–1007.

Schafer, A. J. (1997). *Prosodic parsing: The role of prosody in sentence comprehension*. University of Massachusetts, Amherst.

Schafer, A. J., Carter, J., Clifton, C., & Frazier, L. (1996). Focus in relative clause construal. *Language and Cognitive Processes, 11*(1-2), 135–164.

Schafer, A. J., Speer, S. R., & Warren, P. (2005). Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In M. Tanenhaus & J. Trueswell (Eds.), *Approaches to studying world siutation langauge use: Psycholinguistic, linguistic and computational perspectives on bridging the product and action tradition* (pp. 209–225). Cambridge, MA: MIT Press.

Schmitt, N. (2010). *Researching vocabulary: a vocabulary research manual*. New York: Palgrave Macmillan.

Schmolz, H. (2015). *Anaphora resolution and text retrieval: A linguistic analysis of hypertexts*. Berlin: De Gruyter Mouton.

Schoonen, R., Hulstijn, J., & Bossers, B. (1998). Metacognitive and language-specific knowledge in native and foreign language reading comprehension: An empirical study among Dutch students in grades 6, 8 and 10. *Language Learning, 48*(1), 71–106.

Schreiber, P. A. (1980). On the acquisition of reading fluency. *Journal of Literacy Research, 12*(3), 177–186.

Schreiber, P. A. (1987). Prosody and structure in children's syntactic processing. In Rosalind Horowitz & S. Jay Samuels (Eds.), *Comprehending oral and written*

language (pp. 243–270). New York: Academic Press.

Schreiber, P. A. (1991). Understanding prosody's role in reading acquisition. *Theory Into Practice, 30*(3), 158–164.

Schwanenflugel, P. J., Westmoreland, M. R., & Benjamin, R. G. (2015). Reading fluency skill and the prosodic marking of linguistic focus. *Reading and Writing, 28*, 9–30.

Segalowitz, N., & Hulstijn, J. H. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll, & A. M. B. De Groot (Eds). *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 371–388). Oxford: Oxford University Press.

Sekerina, I. (1997). The Late Closure Principle vs. the Balance Principle: Evidence from on-line processing of ambiguous Russian sentences. In P. Costa (Ed.), *The Proceedings of the Second European Conference on Formal Description of Slavic Languages*. University of Potsdam, Germany.

Sekerina, I. (2004). Cross-linguistic variation in gender use as a parsing constraint: Dutch vs. Russian. Paper presented at the 13th Annual Meeting Formal Description of Slavic Languages, 13 FASL, Columbia, February 24–29.

Sekerina, I., Petrova, K., & Fernández, E. (2003). Relative clause attachment in Bulgarian. Paper presented at the 12th Annual Meeting Formal Description of Slavic Languages, 12 FASL, Ottawa, May 9–11.

Selkirk, E. O. (1981). On the nature of phonological representation. In J. Anderson, J. Laver, & T. Myers (Eds.), *The Cognitive Representation of Speech* (pp. 223–261). Amsterdam: North Holland.

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Selkirk, E. O. (2003). The prosodic structure of function words. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 187–213). Mahwah, NJ: Erlbaum.

Selkirk, E. O., & Tateishi, K. (1988). Constraints on minor phrase formation in Japanese. In G. Larson & D. Brentari (Eds.), *Proceedings of the 24th Annual Meeting of the Chicago Linguistics Society* (pp. 316–336). Chicago: Chicago Linguistics Society.

Selkirk, E. O. (2011). The syntax-phonology interface. In J. Goldsmith, J. Riggle & A. Yu (Eds.), *The Handbook of Phonological Theory* (pp. 435–484). New Jersey: Wiley-Blackwell.

Slabakova, R., White, L., & Brambatti, G. N. (2017). Pronoun Interpretation in the Second Language: Effects of Computational Complexity. *Frontiers in Psychology, 8,* Article 1236.

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, *69*(2), 104–120.

Slowiaczek, L. M. (1990). Effects of lexical stress in auditory word recognition. *Language and Speech*, *33*(1), 47–68.

Slowiaczek, L. M. (1991). Stress and context in auditory word recognition. *Journal of Psycholinguistic Research*, *20*, 465–481.

Slowiaczek, L. M, & Clifton, C. (1980). Subvocalization and reading for meaning. *Journal of Verbal Learning and Verbal Behavior*, *19*(5), 573–582.

Smyth, T. R. & Glencross, D. J. (1986). Information processing deficits in clumsy children. *Australian Journal of Psychology*, *38*, 13–22.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language, 48*(1), 103–130.

Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language*, *58*(2), 574–608.

Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, *25*(2), 249–271.

Speer, S. R., Shih, C., & Slowiaczek, M. L. (1989). Prosodic structure in language understanding: Evidence from tone sandhi in Mandarin. *Language and Speech*, *32*(4), 337– 354.

Sportiche, D. (2013). Binding theory – structure sensitivity of referential dependencies. *Lingua*, *130*, 187–208.

Stabler, E. P. (1994). The finite connectivity of linguistic structure. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 303–336). Lawrence Erlbaum Associates, Inc. (Reprinted in modified form from UCLA Working Papers in Syntax and Semantics, 11, 1992, pp. 87–133)

Stanovich, K. E., Cunningham, A. E., & Feeman, D. J. (1984). Intelligence, cognitive skills, and early reading progress. *Reading Research Quarterly, 19*(3), 278–303.

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language, 86*(1), 142–164.

Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: evidence from event-related brain potentials. *Second Language Research, 25*(1), 13–41.

Sternberg, S. (1975). Memory scanning: New findings and current controversies. *The Quarterly Journal of Experimental Psychology, 27*(1), 1–32 .

Stevenson, R. J., & Vitkovitch, M. (1986). The comprehension of anaphoric relations. *Language and Speech, 29*(4), 335–360.

Stewart, A. J., Pickering, M. J., & Sanford, A. J. (2000). The time course of the influence of implicit causality information: Focusing versus integration accounts. *Journal of Memory and Language, 42*(3), 423–443.

Stirling, L. F., & Huddleston, R. (2002). Deixis and anaphora. In G. K. Huddleston, R. & Pullum (Ed.), *The Cambridge grammar of the English language* (pp. 1449–1565). Cambridge: Cambridge University Press.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language, 48,* 542–562.

Sun, X. (2006). A comparative study on L1 influence and lexical aspect in past tense and progressive marking by Japanese and Chinese EFL learners. *International Journal of Curriculum Development and Practice, 8,* 13-26.

Sun, C. F. (2006). *Chinese: a linguistic introduction*. Cambridge, England: Cambridge University Press.

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*(1), 201–216.

Tan M, Foltz A. (2020). Task sensitivity in L2 English speakers' syntactic processing: evidence for Good-enough processing in self-paced reading. *Front Psychol* (11). doi: 10.3389/fpsyg.2020.575847.

Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia, 56,* 289–301.

Taguchi, E., Gorsuch, G. J., & Sasamoto, E. (2006). Developing second and foreign language reading fluency and its effect on comprehension: A missing link. *The Reading Matrix, 6*(2), 1–18.

Tokizaki, I., (1999). Prosodic phrasing and bare phrase structure. *Proceedings of the North-East Linguistic Society 29,* Vol. 1., 381–395. GLSA, Amherst.

Tokizaki, I., (2001). *Prosodic hierarchy and prosodic boundary*. MS., Sapporo University.

Townsend, J. T., & Ashby, F. G. (1983). *The Stochastic Modeling of Elementary Psychological Processes*. Cambridge: Cambridge University Press.

Traxler, M. J. (2009). A hierarchical linear modeling analysis of working memory and implicit prosody in the resolution of adjunct attachment ambiguity. *J. Psycholinguist Res.* 38, 491–509. doi: 10.1007/s10936-009-9102-x

Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language, 39*(4), 558–592.

Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language, 53*, 204–224.

Tsui, R. K., Tong, X., Fung, L. S. (2016). The role of prosodic reading in English reading comprehension among Cantonese-English bilingual children. *Proceedings of Speech Prosody*, 582–586.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*, 231–270.

Ullman, M. T. (2015). The declarative/procedural model: A neurobiologically motivated theory of first and second language. In B. Van Patten, J. Williams (Eds), *Theories*

in second language acquisition: An introduction. 2nd edition (pp. 135–158). New York: Routledge.

Ullman, M. T., & Pullman, M. Y. (2015). A compensatory role for declarative memory in neurodevelopmental disorders. *Neuroscience & Biobehavioral Reviews, 51,* 205–222.

Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge, and use. In G. Hickok, S. L. Small (Eds), *Neurobiology of language* (pp. 953–968). San Diego, CA: Elsevier.

Ullman, M. T., & Lovelett, J. T. (2018). Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. *Second Language Research*, *34*(1), 39–65.

Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience, 11*(6), 657–671.

Van Berkum, J. J. A., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research, 1146,* 158–171.

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(2), 407–430.

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language, 49*(3), 285–316.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*(2), 157–166.

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language, 65*(3), 247–263.

Van Dyke, J. A., & Johns, C. L. (2012). Memory interference as a determinant of language comprehension. *Language and Linguistic Compass, 6*(4), 193-211.

Van Gelderen, A., Schoonen, R., Stoel, R. D., de Glopper, K., & Hulstijn, J. (2007). Development of adolescent reading comprehension in language 1 and language 2:

A longitudinal analysis of constituent components. *Journal of Educational Psychology*, *99*(3), 477–491.

Van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory and Language*, *55*(3), 335–362.

Vasishth, S., Brüssow, S., Lewis, R. L., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, *32*(4), 685–712.

Verhoeven, L. (1991). Acquisition of biliteracy. *AILA Review*, *8*, 61–74.

Verhoeven, L. (1994). Transfer in bilingual development: The linguistic interdependency hypothesis revisited. *Language Learning*, *44*, 381–415.

Verhoeven, L. (2000). Components of early second language reading and spelling. *Scientific Studies of Reading*, *4*, 313–330.

Verhoeven, L., Perfetti, C., & Pugh, K. (2019). *Dyslexia across languages and writing systems*. Cambridge: Cambridge University Press.

Vigliocco, G., Butterworth, B., & Semenza, C. (1995). Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language*, *34*(2), 186–215.

Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition*, *61*(3), 261–298.

Villata, S., Tabor, W., & Franck, J. (2018). Encoding and retrieval interference in sentence comprehension: Evidence from agreement. *Frontiers in Psychology*, *9*, Article 2. https://doi.org/10.3389/fpsyg.2018.00002

Vonk, W. (1985a). The immediacy of inferences in the understanding of pronouns. In G. Rickheit & H. Strohner (Eds.), *Inferences in text processing* (pp. 205–218). Amsterdam: Elsevier Science Publishers.

Vonk, W. (1985b). On the purpose of reading and the immediacy of processing pronouns. In R. Groner, G. W. McConkie & C. Menz (Eds), *Eye movements and human information processing* (pp. 207-215). New York: North-Holland.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language, 61*(2), 206–237.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes, 25*, 905–945.

Wassenaar, M., Brown, C., & Hagoort, P. (2004). ERP Effects of subject-verb agreement violations in patients with Broca's aphasia. *Journal of cognitive neuroscience, 16*, 553–576.

Watson, D. G., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes, 19*(6), 713–755.

Watson, D., Breen, M., & Gibson, E. (2006). The role of syntactic obligatoriness in the production of intonational boundaries. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(5), 1045–1056.

Weber-Fox, C., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience, 8*(3), 231–256.

Wells, B., Peppé, S., & Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language, 31*(4), 749–778.

Webman-Shafran, R. (2018). Implicit prosody and parsing in silent reading. *J. Res. Read.* 41, 546–563. doi: 10.1111/1467-9817.12124

Wesche, M. B., & Paribakht, T. S. (2010). *Lexical inferencing in a first and second language. Cross-linguistic dimensions*. Bristol: Multilingual Matters.

Witzel, J., Witzel, N., & Nicol, J. (2012). Deeper than shallow: Evidence for structure-based parsing biases in second-language sentence processing. *Applied Psycholinguistics, 33*, 419–456.

Wolters, M., & Byron, D. K. (2000). Prosody and the resolution of pronominal anaphora. *Proceedings of the 18th International Conference on Computational Linguistics*, 919–925. Saarbrücken: Germany, Morgan Kaufmann.

Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain & Language, 108,* 40–55.

Yamashita, J., & Ichikawa, S. (2010). Examining reading fluency in a foreign language: Effects of text segmentation on L2 readers. *Reading in a Foreign Language*, *22*(2), 263–283.

Yao, B., and Scheepers, C. (2018). Direct speech quotations promote low relative-clause attachment in silent reading of English. *Cognition* 176, 248–254. doi: 10.1016/j.cognition.2018.03.017

Yngve, V. H. (1960). A model and a hypothesis for language structure. *Proceedings of the American Philosophical Society, 140*, 444–466.

Yubune, E. (2012). Development of web-based material for chunk-based reading aloud and shadowing. *The Bulletin of the Institute of Human Sciences, Toyo University, 14*, 83–94.

Zhang, Y. & Wang, Y. (2007). Neural plasticity in speech acquisition and learning. *Bilingualism: Language and Cognition*, 10 (2), 147–160.

Zhou, H., Rossi, S., & Chen, B. (2017). Effects of working memory capacity and tasks in processing L2 complex sentence: Evidence from Chinese-English bilinguals. *Frontiers in Psychology, 8,* Article 595.